APPLICATION FOR UNITED STATES LETTERS PATENT

For

METHOD AND APPARATUS FOR PROBABILISTICALLY CLASSIFYING TISSUE IN VITRO AND IN VIVO USING FLUORESCENCE SPECTROSCOPY

By

Rebecca Richards-Kortum

Nirmala Ramanujam

Anita Mahadevan-Jansen

Michele Follen Mitchell

And

Urs Utzinger

METHOD AND APPARATUS FOR PROBABILISTICALLY CLASSIFYING TISSUE IN VITRO AND IN VIVO USING FLUORESCENCE SPECTROSCOPY

Rebecca Richards-Kortum
Nirmala Ramanujam
Anita Mahadevan-Jansen
Michele Follen Mitchell
Urs Utzinger

RELATED APPLICATIONS

This application is a continuing application of U.S. Patent Application Serial No. 08/988,840, filed December 11, 1997 (Attorney Docket No. TUUT:003-1), which is a continuation of U.S. Patent Application Serial No. 403,446, filed March 14, 1995, which issued on December 16, 1997 as U.S. Patent No. 5,697,373; U.S. Patent Application Serial No. 08/666,021, filed July 29, 1999 (Attorney Docket No. TUUT:006-2) which was a Continued Prosecution Application of pending prior application Serial No.08/666,021, filed on June 30, 1998, which is a Continuation Prosecution Application under Rule 53 (37 C.F.R. 1.53 (d)) of Serial No.08/666,021, filed June 19, 1996; and U.S. Patent Application Serial No. 08/693,471, filed June 22, 1999 (Attorney Docket No. TUUT:009--2) which is a Continued Prosecution Application of pending prior application Serial No. 08/693,471, filed on September 8, 1998, which is a continuation application under Rule 53 (37 C.F.R. 1.53 (d)) of Serial No. 08/693,471, filed August 2, 1996, which are all hereby incorporated herein by reference in their entirety.

APPENDICES

Appendices A, B, C and D are included herewith. The Appendices include citations to various references. To the extent that these references provide exemplary experimental details or other information supplementary to that set forth herein, they are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to methods and apparatus of probabilistically classifying tissue *in vivo* and *in vitro* using fluorescence spectroscopy, and more particularly to probabilistically classifying normal, cancerous and precancerous epithelial tissue such as cervical tissue *in vivo* and *in vitro* using fluorescence spectroscopy.

Description of Related Art

Fluorescence, infrared absorption and Raman spectroscopies have been proposed for cancer and precancer diagnosis. Many groups have successfully demonstrated their use in various organ systems. Auto and dye induced fluorescence have shown promise in recognizing atherosclerosis and various types of cancers and precancers. Many groups have demonstrated that autofluorescence may be used for differentiation of normal and abnormal tissues in the human breast and lung, bronchus and gastrointestinal tract. Fluorescence spectroscopic techniques have also been investigated for improved detection of cervical dysplasia.

Although a complete understanding of the quantitative information contained within a tissue fluorescence spectrum has not been achieved, many groups have applied fluorescence spectroscopy for real-time, non-invasive, automated characterization of tissue pathology. Characterization of tissue pathology using auto-fluorescence, *see* Appendix A, References 10-23, as well as photosensitizer induced fluorescence, *see* Appendix A, References 25-27, to discriminate between diseased and non-diseased human tissues *in vitro* and *in vivo* has been described in a variety of tissues. However, these various approaches have not been entirely satisfactory.

Auto-fluorescence spectra of normal tissue, intraepithelial neoplasia and invasive carcinoma have been measured from several organ sites *in vivo*. For example, *in vivo* studies

of the human colon at 370 nm excitation (Appendix A, Reference 13) indicated that a simple algorithm based on fluorescence intensity at two emission wavelengths can be used to differentiate normal colon and adenomatous polyps with a sensitivity and specificity of 100% and 97%, respectively. Shomacker *et al.* (Appendix A, Reference 14) conducted

5 similar studies *in vivo* at 337 nm excitation and demonstrated that a multivariate linear regression algorithm based on laser induced fluorescence spectra can be used to discriminate between normal colon and colonic polyps with a similarly high sensitivity and specificity. Lam *et al.* developed a bronchoscope which illuminates tissue at 442 nm excitation and produces a false color image in near real-time which represents the ratio of fluorescence

10 intensities at 520 nm (green) and 690 nm (red) (Appendix A, References 16 and 17). *In vivo* studies demonstrated that the ratio of red to green auto-fluorescence is greater in normal bronchial tissues than in abnormal bronchial tissues (Appendix A, Reference 16). In a trial with 53 patients, the sensitivity of fluorescence bronchoscopy was found to be 72%, as compared to 50% for conventional white light bronchoscopy (Appendix A, Reference B17).

15 Nonetheless, a reliable diagnostic method and apparatus with improved diagnostic capability for use *in vitro* and *in vivo* is needed to allow faster, more effective patient management and potentially further reduce mortality.

SUMMARY OF THE INVENTION

The present invention advantageously achieves a real time, non-invasive, and

20 automated method and apparatus for classifying normal, cancerous and precancerous tissue in a diagnostically useful manner, such as by histopathological classifications, to allow faster, more effective patient management and potentially further reduce mortality.

One embodiment of the invention is a method of probabilistically classifying a sample of tissue of a mammalian anatomical structure, tissues of which may have various

25 morphological and biochemical states and are classifiable in accordance therewith. The method comprises illuminating the tissue sample with electromagnetic radiation of a wavelength selected to stimulate in the tissues of the mammalian anatomical structure a

- 3 -

fluorescence having spectral characteristics distinguishing between a first plurality of classifications therefor; acquiring fluorescence intensity spectrum sample data for the tissue sample from the illuminating step; obtaining a quantity from fluorescence intensity spectral calibration data, the calibration data being from a calibration set comprising tissues in each one of the first plurality of classifications of a statistically significant set of tissues of the mammalian anatomical structures illuminated with the electromagnetic radiation, and the quantity accounting for a significant amount of variation in the calibration data and showing statistically significant differences between the calibration set tissues in the plurality of classifications; obtaining probability distributions of the calibration data as modified by the quantity for each one of the plurality of classifications; and calculating from the probability distributions and from the sample data as modified by the quantity a probability that the tissue sample belongs in one of the plurality of classifications.

Another embodiment of the invention is a method of probabilistically classifying a sample of tissue of a mammalian anatomical structure, tissues of which may have various morphological and biochemical states and are classifiable in accordance therewith. The method comprises illuminating the tissue sample with electromagnetic radiation of a wavelength selected to stimulate in tissues of the mammalian anatomical structure a fluorescence having spectral characteristics indicative of a classification thereof; detecting a first fluorescence intensity spectrum from the tissue sample resulting from the illuminating step; and calculating a probability that the tissue sample belongs in the classification from a data set comprising the fluorescence intensity spectrum.

A further embodiment of the invention is an apparatus for probabilistically classifying a sample of tissue of a mammalian anatomical structure, tissues of which may have various morphological and biochemical states and are classifiable in accordance therewith. The apparatus comprises a controllable illumination source for generating electromagnetic radiation of a wavelength selected to stimulate in the tissues of the mammalian anatomical structure a fluorescence having spectral characteristics distinguishing between a plurality of classifications therefor; an optical system for illuminating the tissue sample with the electromagnetic radiation and acquiring fluorescence

- 4 -

emissions from the tissue sample; a detector for converting the fluorescence emissions from the tissue sample to intensity spectrum sample data; and a processor coupled to the controllable illumination source for control thereof and coupled to the detector for processing the sample data. The processor comprises means for storing a quantity obtained

5   from fluorescence intensity spectral calibration data, the calibration data being from a calibration set comprising tissues in each one of the first plurality of classifications of a statistically significant set of tissues of the mammalian anatomical structures illuminated with the electromagnetic radiation, and the quantity accounting for a significant amount of variation in the calibration data and showing statistically significant differences between the

10   calibration set tissues in the plurality of classifications; means for storing probability distributions of the calibration data as modified by the first quantity for each one of the plurality of classifications; and means for calculating from the probability distributions and from the sample data as modified by the quantity a probability that the tissue sample belongs in one of the first plurality of classifications.

15   A further embodiment of the invention is a computer program product comprising a computer readable medium having program logic recorded thereon for probabilistically classifying a sample of tissue of a mammalian anatomical structure, tissues of which may have various morphological and biochemical states and are classifiable in accordance therewith. The computer program product comprises means for controlling illumination of

20   the tissue sample with electromagnetic radiation of a wavelength selected to stimulate in the tissues of the mammalian anatomical structure a fluorescence having spectral characteristics distinguishing between a plurality of classifications therefor; means for controlling acquisition of fluorescence intensity spectrum sample data for the tissue sample; a quantity obtained from fluorescence intensity spectral calibration data, the calibration data being

25   from a calibration set comprising tissues in each one of the plurality of classifications of a statistically significant set of tissues of the mammalian anatomical structures illuminated with the electromagnetic radiation, and the quantity accounting for a significant amount of variation in the calibration data and showing statistically significant differences between the calibration set tissues in the plurality of classifications; first probability distributions of the

30   calibration data as modified by the first quantity for each one of the plurality of

classifications; and means for calculating from the probability distributions and from the sample data as modified by the quantity a probability that the tissue sample belongs in one of the plurality of classifications.

BRIEF DESCRIPTION OF THE DRAWINGS

5   Figure 1 is a block diagram of an exemplary fluorescence spectroscopy diagnostic apparatus.

Figures 2A, 2B and 2C are flowcharts of a first exemplary fluorescence spectroscopy diagnostic methods.

Figures 3 and 4 are graphs depicting the performance of the first exemplary 10   fluorescence diagnostic method with 337 nm excitation.

Figures 5A, 5B and 6 are graphs illustrating the performance of the first fluorescence spectrum diagnostic method at 380 nm excitation.

Figures 7A, 7B and 8 are graphs illustrating the performance of the first fluorescence spectrum diagnostic method to distinguish squamous normal tissue from SIL at 460 nm 15   excitation.

Figures 9A, 9B and 10 are graphs illustrating the performance of the first fluorescence spectrum diagnostic method to distinguish low grade SIL from high grade SIL at 460 nm excitation.

Figure 11 is a schematic of the portable fluorimeter used to measure cervical tissue 20   fluorescence spectra at three excitation wavelengths.

Figure 12 is a flow chart of a formal analytical process used to develop the screening and diagnostic algorithms. The text in the dashed-line boxes represent mathematical steps implemented on the spectral data and the text in the solid line boxes represent outputs after each mathematical step (NS - normal squamous, NC - normal columnar, LG - LG SIL and HG - HG SIL).

Figure 13A shows the original spectra, Figure 13B shows the corresponding normalized spectra, and Figure 13C shows the corresponding normalized, mean-scaled spectra at 337 nm excitation from a typical patient.

Figure 14A shows the original spectra, Figure 14B shows the corresponding normalized spectra, and Figure 14C shows the normalized, mean-scaled spectra at 380 nm excitation from the same patient.

Figure 15A shows the original spectra, Figure 15B shows the corresponding normalized spectra, and Figure 15C shows the normalized, mean-scaled spectra at 460 nm excitation from the same patient.

Figure 16 is a plot of the posterior probability of belonging to the SIL category of all SILs and normal squamous epithelia from the calibration set. Evaluation of the misclassified SILs indicates that one samples with CIN III, two with CIN II, two with CIN I and two with HPV are incorrectly classified.

Figure 17 is a plot of the posterior probability of belonging to the SIL category of all SILs and normal columnar epithelia from the calibration data set. Evaluation of the misclassified SILs indicates that three samples with CIN II, three with CIN I and one with HPV are incorrectly classified.

Figure 18 is a plot of the posterior probability of belonging to the HG SIL category of all SILs from the calibration set. Evaluation of the misclassified HG SILs indicates that

three samples with CIN III and three with CIN are incorrectly classified as LG SILs; five samples with CIN I and two with HPV are misclassified as HG SIL.

Figures 19A, 19B and 19C show component loadings (CL) of diagnostic principal components of *constituent* algorithm (1), obtained from normalized spectra at 337, 380 and
5 460 nm excitation, respectively.

Figures 20A, 20B and 20C show component loadings (CL) of diagnostic principal components of *constituent* algorithm (2), obtained from normalized, mean-scaled spectra at 337, 380 and 460 nm excitation, respectively.

Figures 21A, 21B and 21C show component loadings (CL) of diagnostic principal
10 components of *constituent* algorithm (3), obtained from normalized spectra at 337, 380 and 460 nm excitation, respectively.

Figures 22A through 22F illustrate various states of the endocervical canal.

Figures 23 and 24 are graphs showing the optical transmission and excitation emission of cervical mucus.

15 Figure 25 is an exemplary apparatus usable to measure endocervical tissue fluorescence spectra at three excitation wavelengths.

Figure 26 is another exemplary apparatus usable to measure endocervical tissue fluorescence spectra at three excitation wavelengths.

Figures 27 and 28 are graphs showing the optical transmission and excitation
20 emission of fluorinated ethylene-propylene (FEP).

Figures 29, 30 and 31 are exemplary fluorescence spectra obtained from endocervical canal tissue.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

5   Fluorescence spectroscopy has the capability to quickly, non-invasively and quantitatively probe the biochemical and morphological changes that occur as tissue becomes neoplastic. The altered biochemical and morphological state of the neoplastic tissue is reflected in the spectral characteristics of the measured fluorescence. This spectral information can be correlated to tissue histopathology, the current "gold standard" to develop clinically effective screening and diagnostic algorithms. These mathematical

10   algorithms can be implemented in software, thereby enabling automated, fast, non-invasive and accurate pre-cancer screening and diagnosis in the hands of non-experts.

      Specifically, fluorescence spectral data acquired from tissues *in vivo* or *in vitro* is processed in accordance with a multivariate statistical method to achieve the ability to probabilistically classify tissue in a diagnostically useful manner, such as by

15   histopathological classification. Fluorescence occurs when a fraction of the light absorbed by the tissue is re-radiated at emission wavelengths that are longer than the excitation light. Thus, the apparatus includes a controllable illumination device for emitting electromagnetic radiation selected to cause tissue to produce a fluorescence intensity spectrum. Also included are an optical system for applying the plurality of radiation wavelengths to a tissue

20   sample, and a fluorescence intensity spectrum detecting device for detecting an intensity of fluorescence spectra emitted by the sample as a result of illumination by the controllable illumination device. Optionally, the system may include a data processor, connected to the detecting device, for analyzing detected fluorescence spectra to calculate a probability that the sample is abnormal.

Multivariate Statistical Method

The data processor analyzes the detected fluorescence spectra using a multivariate statistical method. The five primary steps involved in the multivariate statistical method are (i) preprocessing of spectral data from each patient to account for inter-patient variation, (ii) partitioning of the preprocessed spectral data from all patients into calibration and prediction sets, (iii) dimension reduction of the preprocessed spectra in the calibration set using principal component analysis, (iv) selection of the diagnostically most useful principal components using a two-sided unpaired student's t-test and (v) development of an optimal classification scheme based on logistic discrimination using the diagnostically useful principal component scores of the calibration set as inputs. These five individual steps of the multivariate statistical method are discussed below in more detail.

Classification of tissue of a specific patient being diagnosed may be performed by including the patient in the prediction set or by applying the diagnostically most useful principal components and a suitable classification scheme specifically to the spectra from the patient's tissue.

(i) Preprocessing. The objective of preprocessing is to calibrate tissue spectra for inter-patient variation which might obscure differences in the spectra of different tissue types. Four methods of preprocessing were invoked on the spectral data: (a) normalization, (b) mean scaling, (c) a combination of normalization and mean scaling and (d) median scaling.

Spectra were normalized by dividing the fluorescence intensity at each emission wavelength by the maximum fluorescence intensity of that sample. Normalizing a fluorescence spectrum removes absolute intensity information; methods developed from normalized fluorescence spectra rely on differences in spectral line shape information for diagnosis. If the contribution of the absolute intensity information is not significant, two advantages are realized by utilizing normalized spectra. First, it is no longer necessary to calibrate for inter-patient variation of normal tissue fluorescence intensity as in the two-

- 10 -

stage method. And second, identification of a colposcopically normal reference site in each patient prior to spectroscopic analysis is no longer needed.

Mean scaling was performed by calculating the mean spectrum for a patient (using all spectra obtained from cervical sites in that patient) and subtracting it from each spectrum

5   in that patient. Mean-scaling can be performed on both unnormalized (original) and normalized spectra. Mean-scaling does not require colposcopy to identify a reference normal site in each patient prior to spectroscopic analysis. However, unlike normalization, mean-scaling displays the differences in the fluorescence spectrum from a particular site with respect to the average spectrum from that patient. Therefore this method can enhance

10   differences in fluorescence spectra between tissue categories most effectively when spectra are acquired from approximately equal numbers of non diseased and diseased sites from each patient.

Median scaling is performed by calculating the median spectrum for a patient (using all spectra obtained from cervical sites in that patient) and subtracting it from each spectrum

15   in that patient. Like mean scaling, median scaling can be performed on both unnormalized (original) and normalized spectra, and median scaling does not require colposcopy to identify a reference normal site in each patient prior to spectroscopic analysis. However, unlike mean scaling, median scaling does not require the acquisition of spectra from equal numbers of non diseased and diseased sites from each patient.

20   *(ii) Calibration and Prediction Data Sets.* The preprocessed spectral data were randomly assigned into either a calibration or prediction set. The multivariate statistical method was developed and optimized using the calibration set.   It was then tested prospectively on the prediction data set.

(iii) Principal Component Analysis. Principal component analysis (PCA) is a linear

25   model which transforms the original variables of a fluorescence emission spectrum into a smaller set of linear combinations of the original variables called principal components that account for most of the variance of the original data set. Principal component analysis is

- 11 -

described in Dillon W.R., Goldstein M., *Multivariate Analysis: Methods and Applications,* John Wiley and Sons, 1984, pp. 23-52, the disclosure of which is expressly incorporated herein by reference. While PCA may not provide direct insight to the morphologic and biochemical basis of tissue spectra, it provides a novel approach of condensing all the spectral information into a few manageable components, with minimal information loss. Furthermore, each principal component can be easily related to the original emission spectrum, thus providing insight into diagnostically useful emission variables.

Prior to PCA, a data matrix is created where each row of the matrix contains the preprocessed fluorescence spectrum of a sample and each column contains the pre-processed fluorescence intensity at each emission wavelength. The data matrix D (RC), consisting of R rows (corresponding to r total samples from all patients in the training set) and C columns (corresponding to intensity at c emission wavelengths) can be written as:

$$D = \begin{pmatrix} D_{11} D_{12} \dots D_{1c} \\ D_{21} D_{22} \dots D_{2c} \\ \\ D_{r1} D_{r2} \dots D_{rc} \end{pmatrix} \qquad (1)$$

The first step in PCA is to calculate the covariance matrix, Z. First, each column of the preprocessed data matrix D is mean-scaled. The mean-scaled preprocessed data matrix, D_m is then multiplied by its transpose and each element of the resulting square matrix is divided by (r-1), where r is the total number of samples. The equation for calculating Z is defined as:

$$Z = \frac{1}{r-1}(D_m{}^t \ D_m) \qquad (2)$$

The square covariance matrix, Z (c x c) is decomposed into its respective eigenvalues and eigenvectors. Because of experimental error, the total number of eigenvalues will always equal the total number of columns (c) in the data matrix D assuming that $c < r$. The goal is

to select n < c eigenvalues that can describe most of the variance of the original data matrix to within experimental error. The variance, V accounted for by the first n eigenvalues can be calculated as follows:

$$V = 100 \left(\frac{\sum_{j=1}^{n} \lambda_j}{\sum_{j=1}^{c} \lambda_j} \right) \tag{3}$$

The criterion used in this analysis was to retain the first n eigenvalues and corresponding eigenvectors that account for 99 % of the variance in the original data set.

Next, the principal component score matrix can be calculated according to the following equation:

$$R = DC \tag{4}$$

where, D (r x c) is the preprocessed data matrix and C (c x n) is a matrix whose columns contain the n eigenvectors which correspond to the first n eigenvalues. Each row of the score matrix R (r x c) corresponds to the principal component scores of a sample and each column corresponds to a principal component. The principal components are mutually orthogonal to each other.

Finally, the component loading is calculated for each principal component. The component loading represents the correlation between the principal component and the variables of the original fluorescence emission spectrum. The component loading can be calculated as shown below:

A: 231018(4Y9601!.DOC)

$$CL_{ij} = \frac{C_{ij}}{\sqrt{S_{ii}}}\sqrt{\lambda_j} \qquad (5)$$

where, CL_{ij} represents the correlation between the ith variable (preprocessed intensity at ith emission wavelength) and the jth principal component. C_{ij} is the ith component of the jth eigenvector, λ_j is the jth eigenvalue and S_{ii} is the variance of the ith variable.

Principal component analysis was performed on each type of preprocessed data matrix, described above. Eigenvalues accounting for 99% of the variance in the original preprocessed data set were retained. The corresponding eigenvectors were then multiplied by the original data matrix to obtain the principal component score matrix R.

(iv) Student's T-Test. Average values of principal component scores were calculated for each histo-pathologic tissue category for each principal component obtained from the preprocessed data matrix. A two-sided unpaired student's t-test was employed to determine the diagnostic contribution of each principal component. Such a test is disclosed in Devore J.L., *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole, 1992, and in Walpole R.E., Myers R.H., *Probability and Statistics for Engineers and Scientists*, Macmillan Publishing Co., 1978, Chapter 7, the disclosures of which are expressly incorporated herein by reference. The hypothesis that the means of the principal component scores of two tissue categories are different were tested for 1) normal squamous epithelia and SILs, 2) columnar normal epithelia and SILs and 3) inflammation and SILs. The t-test was extended a step further to determine if there are any statistically significant differences between the means of the principal component scores of high grade SILs and low grade SILs. Principal components for which the hypothesis stated above were true below the 0.05 level of significance were retained for further analysis.

(v) Logistic Discrimination. Logistic discriminant analysis is a statistical technique that can be used to develop diagnostic methods based on posterior probabilities, overcoming the drawback of the binary decision scheme employed in the two-stage method. This

- 14 -

statistical classification method is based on Bayes theorem and can be used to calculate the posterior probability that an unknown sample belongs to each of the possible tissue categories identified. Logistic discrimination is discussed in Albert A., Harris E.K., *Multivariate Interpretation of Clinical Laboratory Data*, Marcel Dekker, 1987, the

5 disclosure of which is expressly incorporated herein by reference. Classifying the unknown sample into the tissue category for which its posterior probability is highest results in a classification scheme that minimizes the rate of misclassification.

For two diagnostic categories, G_1 and G_2, the posterior probability of being a member of G_1, given measurement x, according to Bayes theorem is:

$$P(G_1 \mid x) = \frac{P(x \mid G_1)\,P(G_1)\,C(2 \mid 1)}{P(x \mid G_1)\,P(G_1)\,C(2 \mid 1) + P(x \mid G_2)\,P(G_2)\,C(1 \mid 2)} \tag{6}$$

10 where $P(x \mid G_i)$ is the conditional joint probability that a tissue sample of type i will have principal component score x, and $P(G_i)$ is the prior probability of finding tissue type i in the sample population. $C(j \mid i)$ is the cost of misclassifying a sample into group j when the actual membership is group i.

The prior probability $P(G_i)$ is an estimate of the likelihood that a sample of type i

15 belongs to a particular group when no information about it is available. If the sample is considered representative of the population, the observed proportions of cases in each group can serve as estimates of the prior probabilities. In a clinical setting, either historical incidence figures appropriate for the patient population can be used to generate prior probabilities, or the practitioner's colposcopic assessment of the likelihood of precancer can

20 be used to estimate prior probabilities.

The conditional probabilities can be developed from the probability distributions of the n principal component scores for each tissue type, i. The probability distributions can be modeled using various techniques. For example, one technique is the gamma function,

which is characterized by two parameters, alpha and beta, which are related to the mean and standard deviation of the data set. The Gamma function is typically used to model skewed distributions and is defined below:

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \tag{7}$$

The gamma function can be used to calculate the conditional probability that a sample from tissue type i, will exhibit the principal component score, x. If more than one principal component is needed to describe a sample population, then the conditional joint probability is simply the product of the conditional probabilities of each principal component (assuming that each principal component is an independent variable) for that sample population.

Another technique is the normal probability density function, *see* Appendix A, Reference 31, which is characterized by μ (mean) and σ (standard deviation).

Use of the multivariate statistical method in four illustrative diagnostic methods is described below in the following four examples.

First Example

Instrumentation

Fluorescence spectra were recorded with a spectroscopic system incorporating a pulsed nitrogen pumped dye laser, an optical fiber probe and an optical multi-channel analyzer at colposcopy. The laser characteristics for the study were: 337, 380 and 460 nm wavelengths, transmitted pulse energy of 50 uJ, a pulse duration of 5 ns and a repetition rate of 30 Hz. The probe includes 2 excitation fibers, one for each wavelength and 5 collection fibers. Rhodamine 6G (8 mg/ml) was used as a standard to calibrate for day to day variations in the detector throughput. The spectra were background subtracted and

- 16 -

normalized to the peak intensity of rhodamine. The spectra were also calibrated for the wavelength dependence of the system.

Figure 1 is an exemplary spectroscopic system for collecting and analyzing fluorescence spectra from cervical tissue. The system incorporates a pulsed nitrogen pumped
5   dye laser 100, an optical fiber probe 101 and an optical multi-channel analyzer 103 utilized to record fluorescence spectra from the intact cervix at colposcopy. The probe 101 comprises a central fiber 104 surrounded by a circular array of six fibers. All seven fibers have the same characteristics (0.22 NA, 200 micron core diameter). Two of the peripheral fibers, 106 and 107, deliver excitation light to the tissue surface; fiber 106 delivers
10   excitation light from the nitrogen laser and fiber 107 delivers light from the dye module (overlap of the illumination area viewed by both optical fibers 106, 107 is greater than 85%). The purpose of the remaining five fibers (104 and 108-111) is to collect the emitted fluorescence from the tissue surface directly illuminated by each excitation fibers 106, 107. A quartz shield 112 is placed at the tip of the probe 101 to provide a substantially fixed
15   distance between the fibers and the tissue surface, so fluorescence intensity can be reported in calibrated units.

Excitation light at 337 nm excitation was focused into the proximal end of excitation fiber 106 to produce a 1 mm diameter spot at the outer face of the shield 112. Excitation light from the dye module 113, coupled into excitation fiber 107 was produced by using
20   appropriate fluorescence dyes; in this example, BBQ (1E-03M in 7 parts toluene and 3 parts ethanol) was used to generate light at 380 nm excitation, and Coumarin 460 (1E-02 M in ethanol) was used to generate light at 460 nm excitation. The average transmitted pulse energy at 337, 380 and 460 nm excitation were 20, 12 and 25 mJ, respectively. The laser characteristics for this example are: a 5 ns pulse duration and a repetition rate of 30 Hz,
25   however other characteristics would also be acceptable. Excitation fluences should remain low enough so that cervical tissue is not vaporized and so that significant photo-bleaching does not occur. In arterial tissue, for example, significant photo-bleaching occurs above excitation fluences of 80 mJ/mm.

The proximal ends of the collection fibers 104, 108-111 are arranged in a circular array and imaged at the entrance slit of a polychromator 114 (Jarrell Ash, Monospec 18) coupled to an intensified 1024-diode array 116 controlled by a multi-channel analyzer 117 (Princeton Instruments, OMA). 370, 400 and 470 nm long pass filters were used to block scattered excitation light at 337, 380 and 460 nm excitation respectively. A 205 ns collection gate, synchronized to the leading edge of the laser pulse using a pulser 118 (Princeton Instruments, PG200), effectively eliminated the effects of the colposcope's white light illumination during fluorescence measurements. Data acquisition and analysis were controlled by computer 119 in accordance with the fluorescence diagnostic method described below in more detail with reference to the flowcharts of Figures 2A-2C.

Method

1. SILs vs. Normal Squamous Tissue at 337 nm excitation. A summary of the fluorescence diagnostic method developed and tested in a previous group of 92 patients (476 sites) is presented here. The spectral data were preprocessed by normalizing each spectrum to a peak intensity of one, followed by mean-scaling. Mean scaling is performed by calculating the mean spectrum for a patient (using all spectra obtained from cervical sites in that patient) and subtracting it from each spectrum in that patient. Next, principal component analysis (PCA) is used to transform the original variables of each preprocessed fluorescence emission spectrum into a smaller set of linear combinations called principal components that account for 99% of the variance of the original data set. Only the diagnostically useful principal components are retained for further analysis. Posterior probabilities for each tissue type are determined for all samples in the data set using calculated prior and conditional joint probabilities. The prior probability is calculated as the percentage of each tissue type in the data. The conditional probability was calculated from the gamma function which modeled the probability distributions of the retained principal components scores for each tissue category. The entire data set was split in two groups: calibration and prediction data set such that their prior probabilities were approximately equal. The method is optimized using the calibration set and then implemented on the prediction set to estimate its performance in an unbiased manner. The methods using PCA

and Bayes theorem were developed using the calibration set consisting of previously collected spectra from 46 patients (239 sites). These methods were then applied to the prediction set (previously collected spectra from another 46 patients; 237 sites) and the current data set of 36 samples.

More specifically, at 337 nm excitation, fluorescence spectra were acquired from a total of 476 sites in 92 patients. The data were randomly assigned to either a calibration set or prediction set with the condition that both sets contain roughly equal number of samples from each histo-pathologic category, as shown in Table 1. Table 1A shows the histo-pathologic classification of samples in the training and the validation set examined at 337 nm excitation, and Table 1B shows the histological classification of cervical samples spectroscopically interrogated *in vivo* from 40 patients at 380 nm excitation and 24 patients in 460 nm excitation.

TABLE 1A

Histology	Training Set	Validation Set
Squamous Normal	127	126
Columnar Normal	25	25
Inflammation	16	16
Low Grade SIL	40	40
High Grade SIL	31	30

TABLE 1B

Histology	380 nm excitation (40 patients)	460 nm excitation (24 patients)
Squamous Normal	82	76
Columnar Normal	20	24
Inflammation	10	11
Low Grade SIL	28	14
High Grade SIL	15	22

- 19 -

The random assignment ensured that not all spectra from a single patient were contained in the same data set. The purpose of the calibration set is to develop and optimize the method and the purpose of the prediction set is to prospectively test its accuracy in an unbiased manner. The two-stage method and the multivariate statistical method were optimized using the calibration set. The performance of these methods were then tested prospectively on the prediction set.

Principal component analysis of mean-scaled normalized spectra at 337 nm excitation from the calibration data set resulted in three principal components accounting for 99% of the total variance. Only, the first two principal components obtained from the preprocessed data matrix containing mean-scaled normalized spectra demonstrate the statistically most significant differences ($P < 0.05$) between normal squamous tissues and SILs (PC1: $P < 1E-25$, PC2: $P < 0.006$). The two-tail P values of the scores of the third principal component were not statistically significant ($P < 0.2$). Therefore, the rest of the analysis was performed using these two principal components. All of the principal components are included in Appendix D.

For excitation at 337 nm, the prior probability was determined by calculating the percentage of each tissue type in the calibration set: 65% normal squamous tissues and 35% SILs. More generally, prior probabilities should be selected to describe the patient population under study; the values used here are appropriate as they describe the prediction set as well.

Posterior probabilities of belonging to each tissue type (normal squamous or SIL) were calculated for all samples in the calibration set, using the known prior probabilities and the conditional probabilities calculated from the gamma function. A cost of misclassification of SILs equal to 0.5 was assumed. Figure 3 illustrates the posterior probability of belonging to the SIL category. The posterior probability is plotted for all samples in the calibration set. This plot indicates that 75% of the high grade SILs have a posterior probability greater than 0.75 and almost 90% of high grade SILs have a posterior probability greater than 0.6. While 85% of low grade SILs have a posterior probability greater than 0.5, only 60% of low grade

SILs have a posterior probability greater than 0.75. More than 80% of normal squamous epithelia have a posterior probability less than 0.25. Note that evaluation of normal columnar epithelia and samples with inflammation using this method results in classifying them as SILs.

5   Figure 4 shows the percentage of normal squamous tissues and SILs correctly classified versus cost of misclassification of SILs for the data from the calibration set. An increase in the SIL misclassification cost results in an increase in the proportion of correctly classified SILs and a decrease in the proportion of correctly classified normal squamous tissues. Note, that varying the cost from .4 to .6 alters the classification accuracy of both

10   SILs and normal tissues by less than 15% indicating that a small change in the cost does not significantly alter the performance of the method. An optimal cost of misclassification would be 0.6-0.7 as this correctly classifies almost 95% of SILs and 80% of normal squamous epithelia, for the prior probabilities used and is not sensitivity to small changes in prior probability.

15   The method was implemented on mean-scaled spectra of the prediction set, to obtain an unbiased estimate of its accuracy. The two eigenvectors obtained from the calibration set were multiplied by the prediction matrix to obtain the new principal component score matrix. Using the same prior probabilities, a cost of misclassification of SILs equal to 0.5, and conditional joint probabilities calculated from the gamma function, all developed from

20   the calibration set, Bayes rule was used to calculate the posterior probabilities for all samples in the prediction set.

Confusion matrices in Tables 2A and 2B show the results of the multivariate statistical method applied to the entire fluorescence emission spectra of squamous normal tissues and SILs at 337 n.m excitation in the calibration set and the prediction set,

25   respectively. A comparison of the sample classification between the prediction and calibration sets indicates that the method performs within 7% on an unknown data set of approximately equal prior probability.

TABLE 2A

Classification	Squamous Normal	Low Grade SIL	High Grade SIL
Squamous Normal	83%	15%	10%
SIL	17%	85%	90%

TABLE 2B

Classification	Squamous Normal	Low Grade SIL	High Grade SIL
Squamous Normal	81%	22%	6%
SIL	19%	78%	94%

5   The utility of another parameter called the component loadings was explored for reducing the number of emission variables required to achieve classification with minimal decrease in predictive ability. Portions of the emission spectrum most highly correlated (correlation > 0.9 or < 0.9) with the component loadings were selected and the reduced data matrix was used to regenerate and evaluate the method. Using intensity at 2 emission

10   wavelengths, the method was developed in an identical manner as was done with the entire emission spectrum. It was optimized using the calibration set and implemented on the prediction set. A comparison of the sample classification based on the method using the entire emission spectrum to that using intensity at 2 emission wavelengths indicates that the latter method performs equally well in classifying normal squamous epithelia and low grade

15   SILs. The performance of the latter method is 6% lower for classifying high grade SILs.

2. SILs vs. Normal Columnar Epithelia and Inflammation at 380 nm Excitation. Principal components obtained from the preprocessed data matrix containing mean-scaled normalized spectra at 380 nm excitation could be used to differentiate SILs from non diseased tissues (normal columnar epithelia and inflammation). The principal components

20   are included in Appendix D. Furthermore, a two-sided unpaired t-test indicated that only

principal component 2 (PC2) and principal component 5 (PC5) demonstrated the statistically most significant differences ($p \leq 0.05$) between SILs and non diseased tissues (normal columnar epithelia and inflammation). The p values of the remaining principal component scores were not statistically significant ($p > 0.13$). Therefore, the rest of the analysis was performed using these two principal components which account collectively for 32% of the variation in the original data set.

Figures 5A and 5B illustrate the measured probability distribution and the best fit of the normal probability density function to PC2 and PC5 of non diseased tissues and SILs, respectively. There is reasonable agreement between the measured and calculated probability distribution, for each case. The prior probability was determined by calculating the percentage of each tissue type in the data set: 41% non diseased tissues and 59% SILs. Posterior probabilities of belonging to each tissue type were calculated for all samples in the data set, using the known prior probabilities and the conditional joint probabilities calculated from the normal probability density function. Figure 6 illustrates the retrospective performance of the diagnostic method on the same data set used to optimize it. The posterior probability of being classified into the SIL category is plotted for all samples evaluated. The results shown are for a cost of misclassification of SILs equal to 50%. Figure 6 indicates that 78% of SILs have a posterior probability greater than 0.5, 78% of normal columnar tissues have a posterior probability less than 0.5 and 60% of samples with inflammation have a posterior probability less than 0.5. Note that, there are only 10 samples with inflammation in this study.

Tables 3A and 3B compare (a) the retrospective performance of the diagnostic method on the data set used to optimize it to (b) a prospective estimate of the method's performance using cross-validation. The method uses mean-scaled normalized spectra at 380 nm excitation to differentiate SILs from non diseased tissues (normal columnar epithelia and inflammation). Table 3A indicates that for a cost of misclassification of 50%, 74% of high grade SILs, 78% of low grade SILs, 78% of normal columnar samples and 60% of samples with inflammation are correctly classified. The unbiased estimate of the method's performance in Table 3B indicates that there is no change in the percentage of correctly

- 23 -

classified SILs and approximately only a 10% decrease in the proportion of correctly classified normal columnar samples.

TABLE 3A

Classification	Normal Columnar	Inflammation	Low Grade SIL	High Grade SIL
Non diseased	78%	60%	21%	26%
SIL	22%	40%	79%	74%

TABLE 3B

Classification	Normal Columnar	Inflammation	Low Grade SIL	High Grade SIL
Non diseased	65%	30%	22%	26%
SIL	35%	70%	78%	74%

3. Squamous Normal Tissue vs. SILs at 460 n.m Excitation. Principal components obtained from the preprocessed data matrix containing mean-scaled normalized spectra at 460 nm excitation could be used to differentiate SIL from normal squamous tissue. These principal components are included in Appendix D. Only principal components 1 and 2 demonstrated the statistically most significant differences ($p \leq 0.05$) between SILs and normal squamous tissues. The p values of the remaining principal component scores, were not statistically significant ($p > 0.06$). Therefore, the rest of the analysis was performed using these two principal components which account collectively for 75% of the variation in the original data set.

Figures 7A and 7B illustrate the measured probability distribution and the best fit of the normal probability density function to PC1 and PC2 of normal squamous tissues and SILs, respectively. There is reasonable agreement between the measured and calculated

- 24 -

probability distribution, for each case. The prior probabilities were determined to be: 67% normal squamous tissues and 33% SILs. Next, posterior probabilities of belonging to each tissue type were calculated for all samples in the data set. Figure 8 illustrates the retrospective performance of the diagnostic method on the same data set used to optimize it. The posterior probability of being classified into the SIL category is plotted for all samples evaluated. The results shown are for a cost of misclassification of SILs equal to 55%. Figure 8 indicates that 92% of SILs have a posterior probability greater than 0.5, and 76% of normal squamous tissues have a posterior probability less than 0.5.

A prospective estimate of the method's performance was obtained using cross-validation. Table 4A and Table 4B compare (a) the retrospective performance of the method on the data set used to optimize it to (b) the prospective estimate of the method's performance using cross-validation. The method uses mean-scaled normalized spectra at 460 nm excitation to differentiate SILs from normal squamous tissues. Table 4A indicates that for a cost of misclassification of SILs equal to 55%, 92% of high grade SILs, 90% of low grade SILs, and 76% of normal squamous samples are correctly classified. The unbiased estimate of the method's performance in Table 4B indicates that there is no change in the percentage of correctly classified high grade SILs or normal squamous tissue; there is a 5% decrease in the proportion of correctly classified low grade SILs.

TABLE 4A

Classification	Normal Squamous	Low Grade SIL	High Grade SIL
Normal Squamous	76%	7%	9%
SIL	24%	93%	91%

TABLE 4B

Classification	Normal Squamous	Low Grade SIL	High Grade SIL
Normal Squamous	75%	14%	9%
SIL	25%	86%	91%

5   *4. Low Grade SILs vs. High Grade SILs at 460 n.m Excitation.* Principal components obtained from the preprocessed data matrix containing normalized spectra at 460 nm excitation could be used to differentiate high grade SILs from low grade SILs. These principal components are included in Appendix D. Principal component 4 (PC4) and principal component 7 (PC7) demonstrated the statistically most significant differences ($p <$

10   0.05) between high grade SILs and low grade SILs. The p values of the remaining principal component scores were not statistically significant ($p > 0.09$). Therefore, the rest of the analysis was performed using these two principal components which account collectively for 8% of the variation in the original data set.

     Figures 9A and 9B illustrate the measured probability distribution and the best fit of

15   the normal probability density function of PC4 and PC7 for normal squamous tissues and SILs, respectively. There is reasonable agreement between the measured and calculated probability distribution, for each case. The prior probability was determined to be: 39% low grade SILs and 61% high grade SILs. Posterior probabilities of belonging to each tissue type

were calculated. Figure 10 illustrates the retrospective performance of the diagnostic method on the same data set used to optimize it. The posterior probability of being classified into the SIL category is plotted for all samples evaluated. The results shown are for a cost of misclassification of SILs equal to 65%. Figure 10 indicates that 82% of high grade SILs have a posterior probability greater than 0.5, and 78% of low grade SILs have a posterior probability less than 0.5.

A prospective estimate of the method's performance was obtained using cross-validation. Table 5A and Table 5B compare (a) the retrospective performance of the method on the data set used to optimize it to (b) the unbiased estimate of the method's performance using cross-validation. The method uses mean-scaled normalized spectra at 460 nm excitation to differentiate high grade from low grade SILs. Table 5A indicates that for a cost of misclassification of 65% 82% of high grade SILs and 78% of low grade SILs are correctly classified. The unbiased estimate of the method's performance in Table 5B indicates that there is a 5% decrease in the percentage of correctly classified high grade SILs and low grade SILs.

TABLE 5A

Classification	Low Grade SIL	High Grade SIL
Low Grade SIL	79%	18%
High Grade SIL	21%	82%

TABLE 5B

Classification	Low Grade SIL	High Grade SIL
Low Grade SIL	72%	27%
High Grade SIL	21%	77%

5   Figures 2A, 2B and 2C are flowcharts of the above-described fluorescence spectroscopy diagnostic methods. In practice, the flowcharts of Figures 2A, 2B and 2C are coded into appropriate form and are loaded into the program memory of computer 119 (Figure 1) which then controls the apparatus of Figure 1 to cause the performance of the diagnostic method.

10   Referring first to Figure 2A, control begin in block 300 where fluorescence spectra are obtained from the patient at 337, 380 and 460 nm excitation. Control then passes to block 301 where the probability of the tissue sample under consideration being SIL is calculated from the spectra obtained from the patient at 337 or 460 nm. This method is shown in more detail with reference to Figure 3B.

15   Control then passes to decision block 302 where the probability of SIL calculated in block 301 is compared against a threshold of 0.5. If the probability is not greater than 0.5, control passes to block 303 where the tissue sample is diagnosed normal, and the routine is ended. On the other hand, if the probability calculated in block 301 is greater than 0.5, control passes to block 304 where the probability of the tissue containing SIL is calculated

based upon the emission spectra obtained from excitation at 380 nm. This method is identical to the method used to calculate probability of SIL from fluorescence spectra due to 337 or 460 nm, and is also presented below in more detail with reference to Figure 3B.

5   Control then passes to decision block 306 where the probability of SIL calculated in block 304 is compared against a threshold of 0.5. If the probability calculated in block 304 is not greater than 0.5, control passes to block 307 where normal tissue is diagnosed and the routine is ended. Otherwise, if decision block 306 determines that the probability calculated in block 304 is greater than 0.5, control passes to block 308 where the probability of high grade SIL is calculated from the fluorescence emission spectra obtained from a 460 nm 10   excitation. This method is discussed below in greater detail with reference to Figure 3C.

Control then passes to decision block 309 where the probability of high grade SIL calculated in block 308 is compared with a threshold of 0.5. If the probability calculated in block 308 is not greater than 0.5, low grade SIL is diagnosed (block 311), otherwise high grade SIL is diagnosed (block 312).

15   Referring now to Figure 2B, the conditioning of the fluorescence spectra by blocks 301 and 304 is presented in more detail. It should be noted that while the processing of block 301 and 304 is identical, block 301 operates on spectra obtained from a 337 or 460 nm excitation, whereas block 304 operates on spectra obtain from a 380 nm excitation. In either case, control begins in block 315 where the fluorescence spectra data matrix, D, is 20   constructed, each row of which corresponds to a sample fluorescence spectrum taken from the patient. Control then passes to block 316 where the mean intensity at each emission wavelength of the detected fluorescence spectra is calculated. Then, in block 317, each spectrum of the data matrix is normalized relative to a maximum of each spectrum. Then, in block 318, each spectrum of the data matrix is mean scaled relative the mean calculated in 25   block 316. The output of block 318 is a preprocessed data matrix, comprising preprocessed spectra for the patient under examination.

A: 231018(4Y96011.DOC)

Control then passes to block 319 where principal component analysis is conducted, as discussed above, with reference to equations 2, 3, 4 and 5. During principal component analysis, the covariance matrix Z (equation (2)), is calculated using a preprocessed data matrix, the rows of which comprise normalized, mean scaled spectra obtained from all patients, including the patient presently under consideration. The result of block 319 is applied to block 321 where a two-sided Student's T-test is conducted, which results in selection of only diagnostic principal components. Control then passes to block 322 where logistic discrimination is conducted, which was discussed above with reference to equations 6 and 7.

The quantity calculated by block 322 is the posterior probability of the sample belonging to the SIL category (block 323)

Referring now to Figure 2C, presented are the details of the determination of the probability of high grade SIL from excitation at 460 nm (block 308, Figure 3A). Control begins in block 324 where the fluorescence spectra data matrix, D, is constructed, each row of which corresponds to a sample fluorescence spectrum taken from the patient. Control then passes to block 326 where each spectrum of the data matrix is normalized relative to a maximum of each spectrum. The output of block 326 is a preprocessed data matrix, comprising preprocessed spectra for the patient under examination. It should be noted that, in contrast to the preprocessing performed in the SIL probability calculating routine of Figure 3B, there is no mean scaling performed when calculating the probability of high grade SIL.

Control then passes to block 327 where principal component analysis is conducted, as discussed above, with reference to equations 2, 3, 4 and 5. During principal component analysis, the covariance matrix Z (equation (2)), is calculated using a preprocessed data matrix, the rows of which comprise normalized, mean scaled spectra obtained from all patients, including the patient presently under consideration. The result of block 327 is applied to block 328 where a two-sided Student's T-test is conducted, which results in selection of only diagnostic principal components. Control then passes to block 329 where

A: 231018(4Y9601!.DOC)

logistic discrimination is conducted, which was discussed above with reference to equations 6 and 7.

The quantity calculated by block 329 is the posterior probability of the sample belonging to the high grade SIL category (block 331).

5   Second Example

The first example described above is limited in two principal ways. A first limitation is that fluorescence spectra were not acquired at all three excitation wavelengths (337, 380 and 460 nm) from every patient in the study. Therefore, analysis of spectral data from these studies did not indicate if the classification accuracy of each of the three *constituent*

10   algorithms developed using spectra at a single excitation wavelength could be improved by utilizing tissue spectra at all three excitation wavelengths. A second limitation of these studies is that the accuracy of *composite* screening and diagnostic algorithms utilizing a combination of the *constituent* algorithms could not be evaluated since tissue spectra were not available at all three excitation wavelengths from the same group of patients.

15   Thus, a first goal of the analysis in this second example is to evaluate the accuracy of constitutient and composite algorithms which address these limitations. Fluorescence spectra acquired *in vivo* at all three excitation wavelengths from 381 cervical sites in 95 patients were analyzed to determine if the accuracy of each of the three *constituent* algorithms previously developed in the analysis of the first example can be improved using tissue

20   spectra at a combination of two or three excitation wavelengths rather than at a single excitation wavelength.

A second goal of the analysis is to integrate the three independently developed *constituent* algorithms which discriminate between pairs of tissue types into *composite* screening and diagnostic algorithms that can achieve discrimination between many of the

25   clinically relevant tissue types. The effective accuracy of a *composite* screening algorithm for the identification of SILs (normal epithelium and inflammation *versus* SIL) and a

composite diagnostic algorithm for the identification of high grade SILs (non-high grade *versus* high grade) was evaluated.

Instrumentation

A schematic of the portable fluorimeter 1 which was used to acquire cervical tissue

5   fluorescence spectra at three excitation wavelengths is shown in Figure 11. The fiber-optic probe 3 includes a central fiber surrounded by a circular array of six fibers; all seven fibers have the same characteristics (0.22 NA, 200 μm core diameter). Three fibers along the diameter of the distal end of the probe (Figure 11) are used for excitation light delivery (overlap of the illumination area viewed by the three excitation fibers is greater than 85%).

10   The purpose of the remaining four fibers is to collect the emitted fluorescence from the area (1 mm diameter) directly illuminated by the probe. A quartz shield 5 at the tip of the distal end of the probe which is in direct tissue contact (Figure 11) provides a fixed distance between the optical fibers and the tissue surface so fluorescence intensity can be measured in calibrated units.

15   Two nitrogen pumped-dye lasers are used to provide illumination at three different excitation wavelengths: one laser serves to deliver excitation light at 337 nm (fundamental) and has a dye module which is used to generate light at 380 nm using the fluorescent dye, BBQ (1E-03 M in 7 parts toluene and 3 parts ethanol). The dye module of the second laser is used to provide illumination at 460 nm, using the fluorescent dye, Coumarin 460 (1E-02

20   M in ethanol). Laser illumination at each excitation wavelength, 337, 380 and 460 nm is coupled into each of the excitation fibers. In this study, the average transmitted pulse energies at 337, 380 and 460 nm excitation were 12, 9 and 14 μJ, respectively. The laser characteristics were a 5 ns pulse duration and a repetition rate of 30 Hz.

The proximal ends of the four emission collection fibers are arranged in a circular

25   array and imaged at the entrance slit of a polychromator coupled to a 1,024 intensified diode array controlled by a multi-channel analyzer. 360, 400 and 470 nm long pass filters are used to block scattered excitation light at 337, 380 and 460 nm excitation, respectively from the

- 32 -

detector. A 205 ns collection gate, synchronized to the leading edge of the laser pulse using a pulser (Princeton Instruments, PG200), eliminates the effects of the colposcope's white light illumination during fluorescence measurements. Data acquisition is computer controlled.

5   *Method*

The method pertains to the development and application of a detection technique for human cervical pre-cancer, both *in vitro* and *in vivo*, based on laser induced fluorescence spectroscopy. Fluorescence spectra from 381 cervical samples in 95 patients were acquired at three excitation wavelengths: 337, 380 and 460 nm. A general multivariate statistical

10   algorithm is then used to analyze and extract clinically useful information from tissue spectra acquired *in vivo*. This experiment includes a screening algorithm to discriminate between SILs and-non SILs (normal squamous and columnar epithelia and inflammation), and a diagnostic algorithm to differentiate high grade SILs from non-high grade SILs (low grade SILs, normal epithelia and inflammation). The retrospective and prospective accuracy

15   of both the screening and diagnostic algorithms is compared to the accuracy of Pap smear screening, *see* Appendix A, Reference 5, and to colposcopy in expert hands, *see* Appendix A, Reference 9.

Clinical measurements. A randomly selected group of non-pregnant patients referred to the colposcopy clinic of the University of Texas MD Anderson Cancer Center on the

20   basis of abnormal cervical cytology was asked to participate in the *in vivo* fluorescence spectroscopy study. Informed consent was obtained from each patient who participated and the study was reviewed and approved by the Institutional Review Boards of the University of Texas, Austin and the University of Texas, MD Anderson Cancer Center. Each patient underwent a complete history and a physical examination including a pelvic exam, a Pap

25   smear and colposcopy of the cervix, vagina and vulva.

After colposcopic examination of the cervix, but before tissue biopsy, fluorescence spectra were acquired on average from two colposcopically abnormal sites, two

colposcopically normal squamous sites and 1 normal columnar site (if colposcopically visible) from each patient. Tissue biopsies were obtained only from abnormal sites identified by colposcopy and subsequently analyzed by the probe to comply with routine patient care procedure. All tissue biopsies were fixed in formalin and submitted for histologic examination. Hemotoxylin and eosin stained sections of each biopsy specimen were evaluated by a panel of four board certified pathologists and a consensus diagnosis was established using the Bethesda classification system; *see* Appendix A, Reference 1. This classification system which has previously been used to grade cytologic specimens has now been extended to classification of histology samples. Samples were classified as normal squamous, normal columnar, inflammation, low grade SIL or high grade SIL. Samples with multiple diagnoses were classified into the most severe histo-pathologic category.

Prior to each patient study, the probe was disinfected and a background spectrum was acquired at all three excitation wavelengths consecutively with the probe dipped in a non-fluorescent bottle containing distilled water. The background spectrum was subtracted from all subsequently acquired spectra at corresponding excitation wavelengths for that patient. Next, with the probe placed on the face of a quartz cuvette containing a solution of Rhodamine 610 dissolved in ethylene glycol (2 mg/L), 50 fluorescence spectra were measured at each excitation wavelength. After calibration, fluorescence spectra were acquired from the cervix: 10 spectra for 10 consecutive pulses were acquired at 337 nm excitation; next, 50 spectra for 50 consecutive laser pulses were measured at 380 nm excitation and then at 460 nm excitation. The data acquisition time was 0.33 s at 337 nm excitation and 1.67 s at each 380 and 460 nm excitation per cervical site. Spectra were collected in the visible region of the electromagnetic spectrum with a resolution of 10 nm (full width at half maximum) and a signal to noise ratio of 30:1 at the fluorescence maximum at each excitation wavelength.

All spectra were corrected for the non-uniform spectral response of the detection system using correction factors obtained by recording the spectrum of an N.I.S.T traceable calibrated tungsten ribbon filament lamp. Spectra from each cervical site at each excitation wavelength were averaged and normalized to the peak fluorescence intensity of the

- 34 -

Rhodamine 610 calibration standard at the corresponding excitation wavelength for that patient; absolute fluorescence intensities are reported in these calibrated units. In this clinical study, fluorescence spectra were acquired at all three excitation wavelengths from each cervical site from a total of 381 sites in 95 patients during colposcopy.

5   *Development of screening and diagnostic algorithms.*   Figure 12 illustrates a schematic of the formal analytical process used to develop screening and diagnostic algorithms for the differential detection of SILs, *in vivo*. In Figure 12, the text in the dashed-line boxes represents the mathematical steps implemented on the spectral data, and the text in the solid-line boxes represent the output after each mathematical process. There are four

10   primary steps involved in the multivariate statistical analysis of tissue spectral data. The first step is to pre-process spectral data to reduce inter-patient and intra-patient variation within a tissue type; the pre-processed spectra are then dimensionally reduced into an informative set of principal components which describe most of the variance of the original spectral data set using Principal Component Analysis (PCA). Next, the principal components which contain

15   diagnostically relevant information are selected using an unpaired, one-sided student's t-test, and finally a classification algorithm based on logistic discrimination is developed using these diagnostically relevant principal components.

In summary, three *constituent* algorithms were developed using multivariate statistical analysis: a *constituent* algorithm (1) that discriminates between SILs and normal

20   squamous tissues, a *constituent* algorithm (2) that discriminates between SILs and normal columnar tissues, and a *constituent* algorithm (3) that differentiates high grade SILs from low grade SILs. The three *constituent* algorithms were then combined to develop two *composite* algorithms: *constituent* algorithms (1) and (2) were combined to develop a *composite* screening algorithm which discriminates between SILs and non SILs; and all

25   three *constituent* algorithms were combined to develop a *composite* diagnostic algorithm which differentiates high grade SILs from non-high grade SILs.

Multivariate statistical analysis of cervical tissue spectra. As a first step, three methods of pre-processing were applied to the spectral data at each excitation wavelength:

1) normalization 2) mean-scaling and 3) a combination of normalization and mean-scaling. Similarly pre-processed spectra at each excitation wavelength were combined to create spectral inputs at the following combinations of excitation wavelengths: (337, 460) nm, (337, 380) nm, (380, 460) nm and (337, 380, 460) nm. Pre-processing of spectral data

5 resulted in four types of spectral inputs (original and three types of pre-processed spectral inputs) at three single excitation wavelengths and at four possible combinations of multiple excitation wavelengths. Hence, there were a total of 12 spectral inputs at single excitation wavelengths and 16 spectral inputs at multiple excitation wavelengths which were evaluated using the multivariate statistical algorithm.

10 Prior to PCA, the input data matrix, D(r x c) was created so each row of the matrix corresponded to the pre-processed fluorescence spectrum of a sample and each column corresponded to the pre-processed fluorescence intensity at each emission wavelength. Spectral inputs at multiple excitation wavelengths were created by arranging spectra at each excitation wavelength in series in the original spectral data matrix. PCA (*see* Appendix A,

15 Reference 28) was used to dimensionally reduce the pre-processed spectral data matrix into a smaller orthogonal set of linear combinations of the emission variables that account for most of the variance of the spectral data set.

Average values of principal component scores were calculated for each principal component of each tissue type. An unpaired, one-sided student's t-test (*see* Appendix A,

20 Reference 29) was employed to determine the diagnostic content of each principal component. The hypothesis that the means of the principal component scores of two tissue types are different was tested for (1) normal squamous epithelia and SILs, (2) normal columnar epithelia and SILs and (3) inflammation and SILs. The t-test was extended a step further to determine if there were any statistically significant differences between the means

25 of the principal component scores of high grade SILs and low grade SILs. Principal components for which the hypothesis stated above was statistically significant ($P < 0.05$) were retained for further analysis.

Next, a statistical classification algorithm was developed using the diagnostically useful principal components to calculate the posterior probability that an unknown sample belongs to each tissue type under consideration. The posterior probability of an unknown sample belonging to each tissue type was calculated using logistic discrimination; *see*

5 Reference 30. The posterior probability is related to the prior and conditional joint probabilities and to the costs of misclassification of the tissue types under consideration. The prior probability of each tissue type was determined by calculating the observed proportion of cases in each group. The cost of misclassification of a particular tissue type was varied from 0 to 1 in 0.1 increments, and the optimal cost was identified when the total number of

10 misclassified samples based on the classification algorithm was a minimum. If there was more than one cost at which the total number of misclassified samples was a minimum, the cost that maximized sensitivity was selected. The conditional joint probabilities were developed by modeling the probability distribution of each principal component of each tissue type using the normal probability density function, *see* Appendix A, Reference 31,

15 which is characterized by μ (mean) and σ (standard deviation). The best fit of the normal probability density function to the probability distribution of each principal component (score) of each tissue type was obtained in the least squares sense, using μ and σ as free parameters of the fit. The normal probability density function was then used to calculate the conditional joint probability that an unknown sample, given that it is from tissue type i, will

20 exhibit a set of principal component scores, x.

The multivariate statistical algorithm was developed and optimized using a calibration set and then tested in an unbiased manner on a prediction set of approximately equal prior probability (Table 6). Data in the prediction set is pre-processed and organized into two prediction datasets in the following way. Spectra obtained from each patient at each

25 excitation wavelength are separately (1) normalized and (2) normalized, followed by mean-scaling. Spectra at each excitation wavelength, processed in a similar manner are concatenated into a vector. Two prediction data matrices are developed. In each matrix, each row is a vector containing similarly pre-processed fluorescence emission spectra at 337, 380 and 460 nm excitation concatenated and each column corresponds to pre-processed

30 fluorescence intensity at a particular excitation emission wavelength pair.

These processed data matrices are then used to test the composite screening algorithm performance. The steps of this test are:

The normalized prediction data matrix (Dn') is multiplied by the reduced eigenvector matrix from normalized spectral data of the calibration set (Cn'). Cn' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 1.

The posterior probabilities that a sample is SIL or normal squamous epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal squamous epithelium and SILs and prior probabilities and optimal costs of misclassification of the calibration set are used.

The normalized, mean-scaled prediction data matrix (Dnm') is multiplied by the reduced eigenvector matrix from normalized, mean-scaled spectral data of the calibration set (Cnm'). Cnm' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 2.

The posterior probabilities that a sample is SIL or normal columnar epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal columnar epithelium and SILs and prior probabilities and optimal costs of misclassification of the calibration set are used.

Using constituent algorithm 1, samples with a posterior probability of being normal squamous epithelium greater than a threshold value are classified as non-SIL. Remaining samples are classified based on the output of constituent algorithm 2. Using constituent algorithm 2, sample with a

- 38 -

posterior probability of being normal columnar epithelium greater than a threshold are classified as non-SIL. The remaining samples are classified as SIL.

The processed data matrices are then used to test the composite diagnostic algorithm performance. The steps of this test are:

The normalized prediction data matrix (Dn') is multiplied by the reduced eigenvector matrix from normalized spectral data of the calibration set (Cn'). Cn' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 1.

The posterior probabilities that a sample is SIL or normal squamous epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal squamous epithelium and SILs and prior probabilities and optimal costs of misclassification of the calibration set are used.

The normalized, mean-scaled prediction data matrix (Dnm') is multiplied by the reduced eigenvector matrix from normalized, mean-scaled spectral data of the calibration set (Cnm'). Cnm' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 2.

The posterior probabilities that a sample is SIL or normal columnar epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal columnar epithelium and SILs and prior probabilities and optimal costs of misclassification of the calibration set are used.

The normalized prediction data matrix (Dn') is multiplied by the reduced eigenvector matrix from normalized spectral data of the calibration set (Cn'). Cn' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 3.

5 The posterior probabilities that a sample is HGSIL or LGSIL are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for HGSILs and LGSILs and prior probabilities and optimal costs of misclassification of the calibration set are used.

10 Using constituent algorithm 1, samples with a posterior probability of being normal squamous epithelium greater than a threshold are classified as non-SIL. Remaining samples are classified based on the output of constituent algorithm 2. Using constituent algorithm 2, sample with a posterior probability of being normal columnar epithelium greater than a threshold are classified as non-SIL. Remaining samples are classified based on the output 15 of constituent algorithm 3. Using constituent algorithm 3, samples with a posterior probability of being LGSIL greater than a threshold are classified as LGSIL. The remaining samples are classified as HGSIL.

The calibration and prediction sets were developed by randomly assigning the spectral data into the two sets with the condition that both contain roughly equal number of 20 samples from each histo-pathologic category. The random assignment ensured that not all spectra from a single patient were contained in the same data set. Table 6 shows the histo-pathologic classification of samples from the calibration and prediction sets. Note that biopsies for histological evaluation were not obtained from colposcopically normal squamous and columnar tissue sites to comply with routine patient care procedure.

TABLE 6

Histo-pathology	Calibration Set	Prediction Set
Normal Squamous	94	94
Normal Columnar	13	14
Inflammation	15	14
Low Grade SIL	23	24
High Grade SIL	35	35

Development of constituent algorithms. The multivariate statistical algorithm was developed and optimized using all 28 types of pre-processed spectral inputs from the calibration set. The algorithm was used to identify spectral inputs which provide the greatest

5 discrimination between the following pairs of tissue types: (1) SILs and normal squamous epithelia, (2) SILs and normal columnar epithelia, (3) SILs and inflammation, and (4) high grade SILs and low grade SILs. The optimal spectral input for differentiating between two particular tissue types was identified when the total number of samples misclassified from the calibration set using the multivariate statistical algorithm was a minimum. The algorithm

10 based on the spectral input that minimized misclassification between the two tissue types under consideration was implemented on the prediction data set.

Three multivariate statistical *constituent* algorithms were developed using tissue spectra at three excitation wavelengths. *Constituent* algorithm (1) was developed to differentiate between SILs and normal squamous epithelia; *constituent* algorithm (2) was

15 developed to differentiate between SILs and normal columnar epithelia; and *constituent* algorithm (3) could be used to discriminate between low grade SILs and high grade SILs.

Development of composite algorithms. Each of the independently developed *constituent* algorithms was intended to discriminate only between pairs of tissue types. A combination of these *constituent* algorithms was required to provide discrimination between

20 several of the clinically relevant tissue types. Therefore, two *composite* algorithms were developed: a *composite* screening algorithm was developed to differentiate between SILs and non SILs (normal squamous and columnar epithelia and inflammation) using *constituent*

- 41 -

algorithms (1) and (2) and a *composite* diagnostic algorithm was developed to differentiate high grade SILs from non-high grade SILs (low grade SILs, normal epithelia and inflammation) using all three *constituent* algorithms.

The *composite* screening algorithm was developed in the following manner. First, *constituent* algorithms (1) and (2) were developed independently using the calibration data set. The classification outputs from both *constituent* algorithms were used to determine if a sample being evaluated is SIL or non-SIL: first, using *constituent* algorithm (1), samples were classified as non SIL if they had a probability that is less than 0.5; otherwise, they were classified as SIL. Next, only samples that were classified as SIL based on the algorithm (1) were tested using algorithm (2). Again, samples were classified as non SIL if their posterior probability was less than 0.5; otherwise they were classified as SIL. The spectral data from the prediction set was evaluated using the *composite* screening algorithm in an identical manner.

The *composite* diagnostic algorithm was implemented in the following manner. The three *constituent* algorithms were developed independently using the calibration set. Algorithms (1) and (2) were implemented on each sample from the calibration data set, as described previously. Only samples that were classified as SIL based on algorithms (1) and (2) were tested using algorithm (3). If samples evaluated using algorithm (3) had a posterior probability greater than 0.5, they were classified as high grade SIL; otherwise they were classified as non-high grade SIL. The spectral data from the prediction set was evaluated using the *composite* diagnostic algorithm in an identical manner.

Results: constituent algorithms (1), (2) and (3). Table 7 summarizes the components of the optimal set of three *constituent* algorithms. Algorithm (1) discriminates between SILs and normal squamous tissues, algorithm (2) discriminates between SILs and normal columnar tissues, and algorithm (3) differentiates high grade (HG) SILs from low grade (LG) SILs. Superscripts in the table refer to the following notes: for the principal component analysis, note 1 - Principal Component, and note 2 - Variance accounted for by principal component; and for logistic discrimination, note 3 - μ (mean) and σ (standard deviation) of

- 42 -

principal component scores of tissue types under consideration, and note 4 - prior probabilities of tissue types under consideration.

TABLE 7

Constituent Algorithms	Excitation Wavelengths	Preprocessing Method	PC[1]	V(%)[2]	$(\mu,\sigma)^{[3]}$	PP[4]
(1) SIL vs. Normal Squamous (NS)	337, 380, 460	normalization	PC1	51	NS: (2.993,1.589) SIL: (2.514,0.671)	NS: 62% SIL: 38%
			PC3	11	NS: (2.631,0.292) SIL: (2.535,0.427)	
			PC7	3	NS: (2.850,0.145) SIL: (2.775,0.209)	
(2) SIL vs. Normal Columnar (NC)	337, 380, 460	normalization mean-scaling	PC1	59	NC: (2.479,0.444) SIL: (2.737,0.482)	NC: 28% SIL: 72%
			PC2	12	NC: (2.894,0.330) SIL: (2.990,0.367)	
			PC4	6	NC: (3.006,0.186) SIL: (3.051,0.167)	
			PC5	3	NC: (3.004,0.101) SIL: (2.994,0.199)	
(3) HG SIL (HG) vs. LG SIL (LG)	337, 380, 460	normalization	PC1	51	LG: (2.755,0.663) HG (2.353,0.759)	LG: 40% HG: 60%
			PC3	11	LG: (2.549,0.394) HG (2.453,0.497)	
			PC6	3	LG: (2.042,0.180) HG (2.100,0.180)	
			PC8	2	LG: (2.486,0.223) HG (2.550,0.130)	

Pre-processing. Figure 13A illustrates average fluorescence spectra per site acquired from cervical sites at 337 nm excitation from a typical patient. All fluorescence intensities are reported in the same set of calibrated units. Corresponding normalized and normalized, mean-scaled spectra are illustrated in Figures 13B and 13C, respectively. Evaluation of the original spectra at 337 nm excitation (Figure 13A) indicates that the fluorescence intensity of SILs is less than that of the corresponding normal squamous tissue and greater than that of the corresponding normal columnar tissue over the entire emission spectrum. Examination of normalized spectra from this patient (Figure 13B) indicates that following normalization, the fluorescence intensity of the normal squamous tissue is greater than that

of corresponding SILs over the wavelength range 360 to 450 nm only; between 460 and 600 nm, the fluorescence intensity of SILs is greater than that of the corresponding normal squamous tissue which in part reflects the longer peak emission wavelength of SILs. A comparison of the spectral line shape of SILs to that of the normal columnar tissue illustrates

5 the opposite phenomenon. The normalized fluorescence intensity of SILs is greater than that of the corresponding normal columnar tissue over the wavelength range 360 to 450 nm; however, between 460 and 600 nm, the fluorescence intensity of the normal columnar tissue is greater than that of the SILs. This spectral difference reflects the longer peak emission wavelength of the normal columnar tissue relative to that of SILs. Further evaluation of

10 normalized spectra in Figure 13B indicates that there are spectral line shape differences between low grade SILs and high grade SILs over the wavelength range 360 to 420 nm.

The corresponding normalized, mean-scaled spectra of this patient, shown in Figure 13C displays differences in the normalized fluorescence spectrum (Figure 13B) from a particular site with respect to the average normalized spectrum from this patient. Evaluation

15 of Figure 13C indicates that between 360 and 450 nm, the normalized, mean-scaled fluorescence intensity of the normal squamous tissue is greater than the mean (Y=0), and that of the normal columnar tissue is less than the mean. Above 460 nm, the opposite phenomenon is observed; the fluorescence intensity of the normal squamous tissue is less than the mean, while that of the normal columnar tissue is greater than the mean. The

20 fluorescence intensity of SILs lies close to the mean and is bounded by the intensities of the two normal tissue types. In addition, between 360 and 420 nm, the normalized, mean-scaled fluorescence intensity of the low grade SIL is slightly greater than the mean, while that of the high grade SIL is less than the mean.

Figure 14A illustrates average fluorescence spectra per site acquired from cervical

25 sites at 380 nm excitation, from the same patient. Figures 14B and 14C show the corresponding normalized, and normalized, mean-scaled spectra, respectively. In Figure 14A, the fluorescence intensity of SILs is less than that of the corresponding normal squamous tissue, with the low grade SIL exhibiting the weakest fluorescence intensity over the entire emission spectrum. Note that the fluorescence intensity of the normal columnar

- 44 -

sample is indistinguishable from that of the low grade SIL. Normalized spectra at 380 nm excitation, (Figure 14B), indicate that over the wavelength range 400 to 450 nm, the fluorescence intensity of the normal squamous tissue is slightly greater than that of SILs and that of the normal columnar tissue is less than that of SILs. The opposite phenomenon is

5 observed above 580 nm. A careful examination of the spectra of the low grade SIL and high grade SIL indicates that between 460 and 580 nm, the normalized fluorescence intensity of the low grade SIL is higher than that of the high grade SIL. The normalized, mean-scaled spectra (Figure 14C) enhances the previously observed normalized spectral line shape differences by displaying them relative to the average normalized spectrum of this patient.

10 Figure 14C indicates that between 400 to 450 nm, the fluorescence intensity of the normal squamous tissue is greater than the mean and that of the normal columnar tissue is less than the mean. The opposite phenomenon is observed above 460 nm. The fluorescence intensity of the SILs is bounded by the intensities of the two normal tissue types over the entire emission spectrum. The low grade SIL and high grade SIL also show spectral line shape

15 differences; above 460 nm, the normalized, mean-scaled fluorescence intensity of the low grade SIL lies above the mean and that of the high grade SIL lies below the mean.

Figures 15A, 15B and 15C illustrate original, normalized and normalized, mean-scaled spectra, respectively at 460 nm excitation from the same patient. Evaluation of Figure 15A indicates that the fluorescence intensity of SILs is less than that of the corresponding

20 normal squamous tissue and greater than that of the corresponding normal columnar sample over the entire emission spectrum. Evaluation of normalized spectra at this excitation wavelength (Figure 15B) demonstrates that below 510 nm, the fluorescence intensity of SILs is less than that of the normal squamous tissue and greater than that of the corresponding normal columnar tissue. Above, 580 nm, the normalized fluorescence

25 intensity of SILs is less than that of the normal columnar tissue and greater then that of normal squamous tissue. Note that there are spectral line shape differences between the low grade SIL and high grade SIL between 580 and 660 nm; the normalized fluorescence intensity of the low grade SIL is greater than that of the high grade SIL. The normalized, mean-scaled spectra shown in Figure 15C reflects the differences observed in the

30 normalized spectra relative to the average normalized spectrum of this patient. Below 510

- 45 -

nm, the fluorescence intensity of the normal squamous tissue is greater than the mean, while that of the normal columnar tissue is less than the mean. Above 580 nm, the opposite phenomenon is observed. The fluorescence intensity of the SILs lies between those of the two normal tissue types. Above 580 nm, the fluorescence intensity of the low grade SIL is

5 greater than the mean and that of the high grade SIL is less than the mean.

Principal Component Analysis and Logistic Discrimination: Constituent algorithm (1) which differentiates SILs from normal squamous tissues. A *constituent* algorithm based on normalized spectra arranged in series at all three excitation wavelengths provided the greatest discrimination between SILs and normal squamous tissues. The algorithm

10 demonstrated an incremental improvement in sensitivity without sacrificing specificity relative to the previously developed *constituent* algorithm (1) that employed normalized, mean-scaled spectra at 337 nm excitation only. Multivariate statistical analysis of normalized tissue spectra at all three excitation wavelengths, indicated three principal components show statistically significant differences between SILs and normal squamous

15 tissues (Table 7). These three principal components account collectively for 65% of the total variance of the spectral data set. Logistic discrimination was used to develop a classification algorithm to discriminate between SILs and normal squamous epithelia based on these three informative principal components. Prior probabilities were determined by calculating the percentage of each tissue type from the data set: 62% normal squamous tissues and 38%

20 SILs. The cost of misclassification of SIL was optimized at 0.7. Posterior probabilities of belonging to each tissue type were calculated for all samples from the data set, using the known prior probabilities, cost of misclassification of SILs and the conditional joint probabilities calculated from the normal probability density function. Figure 16 illustrates the retrospective accuracy of the algorithm applied to the calibration data set. The posterior

25 probability of being classified into the SIL category is plotted for all SILs and normal squamous epithelia. Figure 16 indicates that 92% of high grade SILs and 83% of low grade SILs are correctly classified with a posterior probability greater than 0.5. Approximately 70% of colposcopically normal squamous epithelia are correctly classified with a posterior probability less than 0.5.

A: 231018(4Y9601!.DOC)

The confusion matrix in Table 8 compares the retrospective accuracy of *constituent* algorithm (1) on the calibration data set to its prospective accuracy on the prediction set. In the confusion matrix, the first row corresponds to the histo-pathologic classification and the first column corresponds to the spectroscopic classification of the samples. A prospective

5 evaluation of the algorithm's accuracy indicates that there is a small increase in the proportion of correctly classified low grade SILs and no change in the proportion of correctly classified low grade SILs or normal squamous tissues. Note that the majority of normal columnar tissues and samples with inflammation from both calibration and prediction sets are misclassified as SIL using this algorithm. Evaluation of the misclassified

10 SILs from the calibration set indicates that one sample with CIN III, two with CIN II, two with CIN I and two with HPV are incorrectly classified. From the prediction set, two samples with CIN III, one with CIN II, two with CIN I and one with HPV are incorrectly classified as non-SIL.

TABLE 8

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	68%	8%	7%	17%	9%
SIL	32%	92%	93%	83%	91%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	68%	29%	21%	12%	9%
SIL	32%	71%	79%	88%	91%

15 *Constituent algorithm (2) which differentiates SILs from normal columnar tissues.* The greatest discrimination between SILs and normal columnar epithelia was achieved using a *constituent* algorithm based on normalized, mean-scaled spectra at all three excitation wavelengths. This algorithm demonstrated a substantially improved sensitivity for a similar specificity relative to the previously developed *constituent* algorithm (2) which used

20 normalized, mean-scaled spectra at 380 nm excitation, only. Multivariate statistical analysis of a combination of normalized, mean-scaled tissue spectra at all three excitation wavelengths resulted in four principal components that demonstrate statistically significant differences between SILs and normal columnar epithelia (Table 7). These four principal

components collectively account for 80% of the total variance of the spectral data set. Logistic discrimination was employed to develop a classification algorithm to discriminate between SILs and normal columnar epithelia. The prior probabilities were determined to be: 28% normal columnar tissues and 72% SILs. The optimized cost of misclassification of SIL

5   was equal to 0.58. Posterior probabilities of belonging to each tissue type were calculated for all samples from the data set. Figure 17 illustrates the retrospective accuracy of the algorithm applied to the calibration data set. The posterior probability of being classified into the SIL category is plotted for all SILs and normal columnar samples examined. Figure 17 graphically indicates that 91% of high grade SILs and 83% of low grade SILs have a

10   posterior probability that is greater than 0.5. 76% of colposcopically normal columnar epithelia are correctly classified with a posterior probability less than 0.5.

The confusion matrix in Table 9 compares the retrospective accuracy of *constituent* algorithm (2) on the calibration data set to its prospective accuracy on the prediction set. The first column corresponds to the spectroscopic classification and the first row corresponds to

15   the histo-pathologic classification. The prospective accuracy of the algorithm (Table 9) indicates that there is a small increase in the proportion of correctly classified low grade SILs and a small decrease in the proportion of correctly classified high grade SILs; there is approximately a 10% decrease in the proportion of correctly classified normal columnar tissues. Note that the majority of normal squamous tissues and samples with inflammation

20   from both the calibration and prediction sets are misclassified as SIL using this algorithm. Evaluation of the misclassified SILs from the calibration set indicates that three samples with CIN II, three with CIN I and one with HPV are incorrectly classified. From the prediction set, two samples with CIN III, three with CIN II, and three with CIN I are incorrectly classified.

TABLE 9

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	7%	77%	27%	17%	9%
SIL	93%	23%	73%	83%	91%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	5%	64%	27%	13%	14%
SIL	95%	36%	73%	87%	86%

Constituent algorithm (3) which differentiates High Grade SILs and Low Grade SILs. A combination of normalized spectra at all three excitation wavelengths significantly enhanced the accuracy of the previously developed *constituent* algorithm (3) which

5   differentiated high SILs from low grade SILs using normalized spectra at 460 nm excitation. Multivariate statistical analysis of normalized spectra at all three excitation wavelengths resulted in four statistically significant principal components, that account collectively for 67% of the total variance of the spectral data set (Table 7). Again, a probability based classification algorithm was developed to differentiate high grade SILs from low grade SILs.

10   The prior probability was: 40% low grade SILs and 60% high grade SILs. The optimal cost of misclassification of high grade SIL was equal to 0.51. Posterior probabilities of belonging to each tissue type were calculated. Figure 18 illustrates the retrospective accuracy of the algorithm applied to the calibration data set. The posterior probability of being classified into the high grade SIL category is plotted for all SILs evaluated. Figure 18 indicates that

15   83% of high grade SILs have a posterior probability greater than 0.5, and 70% of low grade SILs have a posterior probability less than 0.5.

The confusion matrix in Table 10 compares the retrospective accuracy of *constituent* algorithm (3) on the calibration set to its prospective accuracy on the prediction set. The first column corresponds to the spectroscopic classification and the first row corresponds to the

20   histo-pathologic classification. Its prospective accuracy indicates that there is a 5% decrease in the proportion of correctly classified low grade SILs and no change in the proportion of correctly classified high grade SILs. From the calibration set, six high grade SILs are misclassified; three samples with CIN III and three with CIN II are misclassified as low

A: 231018(4Y9601!.DOC)

grade SIL. The misclassified low grade SILs comprise of five samples with CIN I and two with HPV. From the prediction set, five high grade SILs are misclassified; two have CIN III and three have CIN II. Of the ten misclassified low grade SILs from the prediction set, seven have CIN I and three have HPV.

5

TABLE 10

Classification in Calibration Set	LG SIL	HG SIL
LG SIL	69%	17%
HG SIL	31%	83%
Classification in Prediction Set	LG SIL	HG SIL
LG SIL	63%	19%
HG SIL	37%	81%

"Full-parameter" composite screening and diagnostic algorithms. A *composite* screening algorithm was developed to differentiate SILs and non-SILs (normal squamous and columnar epithelia and inflammation) and a *composite* diagnostic algorithm was developed to differentiate high grade SILs from non-high grade SILs (low grade SILs, normal epithelia and inflammation). The effective accuracy of both *composite* algorithms were compared to those of the *constituent* algorithms from which they were developed and to the accuracy of current detection modalities; *see* Appendix A, References 5 and 9.

A composite screening algorithm which discriminates between SILs and non SILs. A *composite* screening algorithm to differentiate SILs from non-SILs was developed using a combination of the two *constituent* algorithms: algorithm (1) which differentiates SILs from normal squamous tissues and algorithm (2) which differentiates SILs from normal columnar epithelia. The optimal cost of misclassification of SIL was equal to 0.66 for *constituent* algorithm (1) and 0.64 for *constituent* algorithm (2). Only the costs of misclassification of SIL of the two *constituent* algorithms was altered for the development of the *composite* screening algorithm. These costs were selected to minimize the total number of misclassified samples.

- 50 -

The accuracy of the *composite* screening algorithm on the calibration and prediction data sets is illustrated in the confusion matrix in Table 11. The first column corresponds to the spectroscopic classification and the first row corresponds to the histo-pathologic classification. Examination of the confusion matrix indicates that the algorithm correctly

5   classifies approximately 90% of high grade SILs and 75% of low grade SILs from the calibration data set. Furthermore, approximately, 80% of normal squamous tissues and 70% of normal columnar epithelia from the calibration set are correctly classified. Evaluation of the prediction set indicates that there is a small change in the proportion of correctly classified high grade SILs and low grade SILs. There is a negligible change in the correct

10   classification of normal squamous and columnar tissues. Note that while 80% of samples with inflammation from the calibration set are incorrectly classified as SIL, only 43% of these samples from the prediction set are incorrectly classified.

TABLE 11

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	79%	69%	20%	26%	11%
SIL	21%	31%	80%	74%	89%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	75%	69%	57%	25%	14%
SIL	25%	31%	43%	75%	86%

A comparison of the accuracy of the *composite* screening algorithm (Table 11) to

15   that of each of the *constituent* algorithms (1) (Table 8) and (2) (Table 9) on the same spectral data set indicates that in general, there is less than a 10% decrease in the proportion of correctly classified SILs using the *composite* screening algorithm relative to using either of the *constituent* algorithms independently. Note, however, that the proportion of correctly classified normal (squamous and columnar) epithelia is substantially higher using the

20   *composite* algorithm relative to using either of the *constituent* algorithms independently. These results confirm that utilization of a combination of the two *constituent* algorithms,

- 51 -

significantly reduces the false-positive rate relative to that using each algorithm independently. Evaluation of the spectroscopically misclassified SILs from the calibration set (Table 6) indicates that only one sample with CIN III, three with CIN II, two with CIN I and four with HPV are incorrectly classified. From the prediction data set (Table 6), two samples with CIN III, four with CIN II, three with CIN I and one sample with HPV are incorrectly classified.

A composite diagnostic algorithm which differentiates High Grade SILs from non-High Grade SILs. A *composite* diagnostic algorithm which differentially detects high grade SILs was developed using a combination of all three *constituent* algorithms: algorithm (1) which differentiates SILs from normal squamous tissues, algorithm (2) which differentiates SILs from normal columnar epithelia, and algorithm (3) which differentiates high grade SILs from low grade SILs. The optimal costs of misclassification of SIL was equal to 0.87 for algorithm (1) and 0.65 for algorithm (2); the optimal cost of misclassification of high grade SIL was equal to 0.49 for algorithm (3). Only the costs of misclassification of SIL of *constituent* algorithms (1) and (2) and the cost of misclassification of high grade SIL of *constituent* algorithm (3) were altered during development of the *composite* diagnostic algorithm. These costs were selected to minimize the total number of misclassified samples.

The results of the *composite* diagnostic algorithm on the calibration and prediction sets are shown in the confusion matrix in Table 12. The first column corresponds to the spectroscopic classification and the first row corresponds to the histo-pathologic classification. The algorithm correctly classifies 80% of high grade SILs, 74% of low grade SILs and more than 80% of normal epithelia. Evaluation of the prediction set using this *composite* algorithm indicates that there is only a 3% decrease in the proportion of correctly classified high grade SILs and a 7% decrease in the proportion of correctly classified low grade SILs. There is less than a 10% decrease in the proportion of correctly classified normal epithelia. A comparison between the calibration and prediction sets indicates that while more than 70% of samples with inflammation from the calibration data set are incorrectly classified as high grade SIL, only 14% of samples with inflammation from the prediction set are incorrectly identified. Due to the relatively small number of samples

examined in this histo-pathologic category, the results presented here do not conclusively establish if the algorithm is capable of correctly identifying inflammation.

TABLE 12

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non HG SIL	84%	77%	27%	74%	20%
HG SIL	16%	23%	73%	26%	80%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non HG SIL	85%	69%	86%	67%	23%
HG SIL	15%	31%	14%	33%	77%

A comparison of the accuracy of the *composite* diagnostic algorithm to that of *constituent* algorithm (3) which differentiates high grade SILs from low grade SILs (Table 10) indicates there is less than a 5% decrease in the proportion of correctly classified high grade SILs and a 5% increase in the proportion of correctly classified low grade SILs using the *composite* diagnostic algorithm relative to using the *constituent* algorithm (3). Evaluation of the high grade SILs from the calibration set (Table 12) that were incorrectly classified indicates that three samples with CIN III and four with CIN II are incorrectly classified. From the prediction set, four samples with CIN III and five with CIN II are incorrectly classified.

Third Example

A goal of the analysis in this third example is to determine if fluorescence intensities at a reduced number of excitation-emission wavelength pairs can be used to re-develop *constituent* and *composite* algorithms that can achieve classification with a minimum decrease in predictive ability. A significant reduction in the number of required fluorescence excitation-emission wavelength pairs could enhance the development of a cost-effective clinical fluorimeter. The accuracy of the *constituent* and *composite* algorithms based on the reduced emission variables was compared to the accuracy of those that utilize entire fluorescence emission spectra.

- 53 -

Instrumentation

The fluorescence emission spectra obtained with the instrumentation of the Second Example were used to demonstrate the method of this Third Example.

Method

5   *"Reduced-parameter" composite screening and diagnostic algorithms: Component Loadings.* A component loading represents the correlation between each principal component and the original pre-processed fluorescence emission spectra at a particular excitation wavelength. Figures 19A, 19B and 19C illustrate component loadings of the diagnostically relevant principal components of *constituent* algorithm (1) obtained from

10   normalized spectra at 337, 380 and 460 nm excitation, respectively. Figures 20A, 20B and 20C display component loadings that correspond to the diagnostically relevant principal components of *constituent* algorithm (2) obtained from normalized, mean-scaled spectra at 337, 380 and 460 nm excitation, respectively. Finally, Figures 21A, 21B and 21C display the component loadings corresponding to the diagnostically relevant principal components of

15   *constituent* algorithm (3), obtained from normalized spectra at 337, 380 and 460 nm excitation, respectively. In each graph shown, the abscissa corresponds to the emission wavelength range at a particular excitation wavelength and the ordinate corresponds to the correlation coefficient of the component loading. Correlation coefficients of the component loading above 0.5 and below -0.5 are considered to be significant.

20   Figures 19A, 20A and 21A display component loadings of principal components of *constituent* algorithms (1), (2) and (3), respectively, obtained from pre-processed spectra at 337 nm excitation. A closer examination indicates that component loading 1 is nearly identical for all three algorithms. Evaluation of this loading indicates that it is positively correlated with corresponding emission spectra over the wavelength range 360-440 nm and

25   negatively correlated with corresponding emission spectra over the wavelength range 460-660 nm. All remaining principal components of all three algorithms display a correlation

- 54 -

between -0.5 and 0.5, except component loading 4 of algorithm (2) (Figure 20A) which displays a positive correlation of 0.75 with the corresponding emission spectra at 460 nm.

Figures 19B, 20B and 21B display component loadings that correspond to the diagnostically relevant principal components of *constituent* algorithms (1), (2) and (3),
5   respectively obtained from pre-processed spectra at 380 nm excitation. Component loading 1 of all three algorithms is positively correlated with corresponding emission spectra over the wavelength range, 400-450 nm. Between 500-600 nm, only component loading 1 of algorithm (2) (Figure 20B) is correlated negatively with corresponding emission spectra. However, examination of component loading 3 of algorithm (1) (Figure 19B) and algorithm
10   (3) (Figure 21B) indicates that they are also negatively correlated with corresponding emission spectra from 500-600 nm. Only component loading 2 of algorithm (2) (Figure 20B) is positively correlated with corresponding emission spectra from 500-600 nm. Also note that component loading 3 of algorithm (1) (Figure 19B) and component loadings 3 and 6 of algorithm (3) (Figure 21B) display a positive correlation with corresponding emission
15   spectra at approximately 640 nm.

Figures 19C, 20C and 21C display component loadings that correspond to the diagnostic principal components of *constituent* algorithms (1), (2) and (3), respectively obtained from pre-processed spectra at 460 nm excitation. Note that only component loading 1 displays a negative correlation (< -0.5) with corresponding emission spectra for all three
20   algorithms. This component loading is correlated with corresponding emission spectra over the wavelength range 580-660 nm. The remaining principal components of all three algorithms display a correlation between -0.5 and 0.5.

The component loadings at all three excitation wavelengths of all three *constituent* algorithms were evaluated to select fluorescence intensities at a minimum number of
25   excitation-emission wavelength pairs required for the previously developed *constituent* and *composite* algorithms to perform with a minimal decrease in classification accuracy. Portions of the component loadings of the three *constituent* algorithms most highly correlated (correlation > 0.5 or < -0.5) with corresponding emission spectra at each

excitation wavelength were selected and the reduced data matrix was then used to regenerate and evaluate the *constituent* and *composite* algorithms. It was iteratively determined that fluorescence intensities at a minimum of 15 excitation-emission wavelength pairs are required to re-develop *constituent* and *composite* algorithms that demonstrate a minimum

5   decrease in classification accuracy. At 337 nm excitation, fluorescence intensities at two emission wavelengths between 360-450 nm and intensities at two emission wavelengths between 460-660 nm were selected. At 380 nm excitation, intensities at two emission wavelengths between 400-450 nm and intensities at four emission wavelengths between 500-640 nm were selected. Finally, at 460 nm excitation, fluorescence intensities at five

10   emission wavelengths over the range 580-660 nm was selected.

Table 13A lists 18 excitation-emission wavelength pairs needed to re-develop the three *constituent* algorithms (1), (2) and (3) with a minimal decrease in classification accuracy. These excitation-emission wavelength pairs are also indicated on the component loading plots in Figures 19, 20 and 21. The bandwidth at each emission wavelength is 10

15   nm.

TABLE 13A

Algorithm (1)	Algorithm (2)	Algorithm (3)
$(\lambda_{exc}, \lambda_{emm})$	$(\lambda_{exc}, \lambda_{emm})$	$(\lambda_{exc}, \lambda_{emm})$
337, 410 nm	337, 410 nm	337, 410 nm
337, 430 nm	337, 430 nm	337, 430 nm
337, 460 nm	337, 460 nm	337, 460 nm
337, 510 nm	337, 510 nm	337, 510 nm
337, 580 nm	337, 580 nm	337, 580 nm
380, 410 nm	380, 410 nm	380, 410 nm
380, 430 nm	380, 430 nm	380, 430 nm
380, 460 nm	380, 460 nm	380, 460 nm
380, 510 nm	380, 510 nm	380, 510 nm
380, 580 nm	380, 580 nm	380, 580 nm
380, 640 nm	380, 600 nm	380, 640 nm
460, 510 nm	460, 510 nm	460, 510 nm
460, 580 nm	460, 580 nm	460, 580 nm
460, 600 nm	460, 600 nm	460, 600 nm
460, 620 nm	460, 620 nm	460, 620 nm
460, 640 nm	460, 660 nm	460, 640 nm

Reduced-parameter composite algorithms. Using the fluorescence intensities only at the selected excitation-emission wavelength pairs, the three *constituent* algorithms were re-developed using the same formal analytical process as was done previously using the entire fluorescence emission spectra at all three excitation wavelengths (Figure 12). The three *constituent* algorithms were then independently optimized using the calibration set and tested prospectively on the prediction data set. They were combined as described previously into *composite* screening and diagnostic algorithms. The effective accuracy of these reduced-parameter *composite* algorithms were compared to that of the full-parameter *composite* algorithms developed previously using fluorescence emission spectra at all three excitation wavelengths.

Table 13B contains fluorescence intensities at 15 of the previous 18 excitation-emission wavelength pairs needed to redevelop the three constituent algorithms with a minimal decrease in classification accuracy. This table indicates that three variables are eliminated and the bandwidths of intensities at four excitation-emission wavelength pairs are increased by approximately a factor of four. These results establish that a further reduction

- 57 -

in the number of emission variables and an increase in bandwidth minimally affect the classification accuracy of the algorithms. The benefit of eliminating the three emission variables and increasing the bandwidth of four emission variables is that it can reduce the total integration time needed to measure the fluorescence parameters from the tissue.

5

TABLE 13B

Excitation, Emission	Old Bandwidth (nm)	New Bandwidth (nm)
337 nm, 410 nm	10	80
337 nm, 430 nm	10	Eliminated
337 nm, 460 nm	10	20
337 nm, 510 nm	10	60
337 nm, 580 nm	10	60
380 nm, 410 nm	10	Eliminated
380 nm, 430 nm	10	Eliminated
380 nm, 510 nm	10	60
380 nm, 460 nm	10	20
380 nm, 580 nm	10	10
380 nm, 600 nm	10	10
380 nm, 640 nm	10	10
460 nm, 510 nm	10	10
460 nm, 580 nm	10	10
460 nm, 600 nm	10	10
460 nm, 620 nm	10	10
460 nm, 640 nm	10	10
460 nm, 660 nm	10	10

Table 14 displays the accuracy of the reduced-parameter *composite* screening algorithm (based on fluorescence intensities at 15 excitation-emission wavelength pairs) which discriminates between SILs and non-SILs applied to the calibration and prediction sets. The first column corresponds to the spectroscopic classification and the first row corresponds to the histo-pathologic classification. A comparison between the calibration and prediction data sets indicates that there is less than a 10% decrease in the proportion of correctly classified SILs and normal squamous tissues from the prediction set. Note however that there is a 20% increase in the proportion of correctly classified normal columnar

A: 231018(4Y9601!.DOC)

epithelia and a 40% increase in the proportion of correctly classified samples with inflammation from the prediction set.

TABLE 14

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	73%	46%	13%	17%	15%
SIL	27%	54%	87%	83%	85%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non SIL	72%	64%	50%	25%	11%
SIL	28%	36%	50%	75%	89%

The accuracy of the reduced-parameter *composite* screening algorithm (Table 14) was compared to that of the full-parameter *composite* screening algorithm (Table 11) applied to the same spectral data set. A comparison indicates that in general there is less than a 10% decrease in the accuracy of the reduced-parameter *composite* algorithm relative to that of the full-parameter *composite* screening algorithm, except for a 20% decrease in the proportion of correctly classified normal columnar epithelia from the calibration set tested using the reduced-parameter *composite* screening algorithm (Table 14).

Table 15 displays the accuracy of the reduced-parameter *composite* diagnostic algorithm that differentially identifies high grade SILs from the calibration and prediction sets. The first column corresponds to the spectroscopic classification and the first row corresponds to the histo-pathologic classification. A comparison of sample classification between the calibration and prediction data sets indicates that there is negligible change in the proportion of correctly classified high grade, low grade SILs and normal squamous epithelia. Note that there is approximately a 20% increase in the proportion of correctly classified normal columnar epithelia and samples with inflammation from the prediction set.

TABLE 15

Classification in Calibration Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non HG SIL	79%	62%	40%	65%	23%
HG SIL	21%	38%	60%	35%	77%
Classification in Prediction Set	Normal Squamous	Normal Columnar	Inflammation	LG SIL	HG SIL
Non HG SIL	82%	86%	64%	63%	20%
HG SIL	18%	14%	36%	37%	80%

A comparison of the *composite* diagnostic algorithm based on the reduced emission variables (Table 15) to that using fluorescence emission spectra at all three excitation wavelengths (Table 12) applied to the same spectral data set indicates that in general, the

5   accuracy of the reduced-parameter *composite* diagnostic algorithm is within 10% of that reported for the full-parameter *composite* diagnostic algorithm. However, a comparison between Tables 12 and 15 indicates that there is approximately a 15% decrease and a 20% increase in the proportion of correctly classified normal columnar epithelia from the calibration and prediction sets (Table 15), respectively which were tested using the reduced-

10   parameter *composite* diagnostic algorithm. The opposite trend is observed for samples with inflammation tested using the reduced-parameter *composite* diagnostic algorithm (Table 15).

Table 16 compares the sensitivity and specificity of the full-parameter and reduced-parameter *composite* algorithms to that of Pap smear screening, *see* Appendix A, Reference 5, and colposcopy in expert hands, *see* Appendix A, Reference 9. Table 16 indicates that the

15   *composite* screening algorithms have a similar specificity and a significantly improved sensitivity relative to Pap smear screening. A comparison of the sensitivity of the *composite* screening algorithms to that of colposcopy in expert hands for differentiating SILs from non SILs indicates that these algorithms demonstrate a 10% decrease in sensitivity, but a 20% improvement in specificity. The *composite* diagnostic algorithms and colposcopy in expert

20   hands discriminate high grade SILs from non-high grade SILs with a very similar sensitivity and specificity. A comparison between the full-parameter and reduced-parameter *composite* algorithms indicates that the algorithms based on the reduced emission variables

A: 231018(4Y96011.DOC)

demonstrate a similar classification accuracy relative to those that employ fluorescence emission spectra at all three excitation wavelengths.

TABLE 16

	SILs vs. NON SILs		HG SIL vs. Non HG SIL	
Classification	Sensitivity	Specificity	Sensitivity	Specificity
Pap Smear	62%±23	68%±21	N/A	N/A
Colposcopy in Expert Hands	94%±6	48%±23	79%±23	76%±13
Original Composite Algorithm	82%±1.4	68%±0.0	79%±2	78%±6
Reduced Composite Algorithm	84%±1.5	65%±2	78%±0.7	74%±2

5

Fourth Example

Instrumentation and methods suitable for characterizing tissue of epithelial lined viscus including, for example, the endocervical canal, are now described. It is known that atypical colposcopic tissue patterns occur with some frequency at the transformation zone
10   between the squamous and columnar epithelium in the endocervical canal; *see* Burke L, Antonioli DA and Ducatman BS. Colposcopy, Text and Atlas, pp. 47, 48, 61 and 62, Appleton and Large, Norwalk CT (1991). In many women, this transformation zone (also known as the squamocolumnar junction) is located well within the endocervical canal and is not easily subjected to colposcopy or fluorescence spectroscopy with systems that are
15   intended primarily to assess the ectocervix. In addition, cervical lesions that exist on the ectocervix often extend into the endocervical canal, and characterization of the lesion within the endocervical canal is often an important matter. It is therefore desirable to provide a means to subject the endocervical canal, including the transformation zone, to fluorescence spectroscopy.

20   Referring now to Figures 22A through 22F, shown are simplified representations of the cross section of the os of the endocervical canal and surrounding tissue illustrating the locations of the squamous epithelium (SE), columnar epithelium (CE) and transformation zone (TZ) of the uterus at various stages of maturity and for various medical conditions.

Specifically. Figure 22A shows the neonate uterus, Figure 13B shows the premenarchal uterus, Figure 22C shows the menarchal uterus, Figure 22D shows the menstruating uterus, Figure 22E shows the menopausal uterus, and Figure 22F shows the postmenopausal uterus. As can be seen, the transformation zone TZ can appear on the ectocervix (for example, menstruating, Figure 3D), or well within the edocervical canal (for example, postmenopausal, Figure 3F), or anywhere in between. Since the most common location for CIN and metaplasia is at or near the transformation zone, it is critical that the transformation zone be imaged when conducting fluorescence spectroscopy. This is of particular importance in menopause and postmenopause because most cervical carcinomas occur at this age, and this is when the transformation zone is most deeply within the endocervical canal.

Other general observations of the morphology of the endocervical canal are worthy of note. After the external os, which follows a funnel type opening, the endocervical canal enlarges and gets smaller again at the inner os. The uterus opens to its full size after the internal os by a small angle. The canal can be filled inside with non-neoplastic additional tissue like polyps and synechia. Polyps may fill the canal. Atrophy may be present, which results in an abnormal form of the wall (missing folds). In addition, it is known that stenosis may occur after LEEP treatments.

The folds of the columnar epithelium may typically be several centimeters deep with varying shapes. For example, in one uterus that was studied after removal by hysterectomy, the folds were a maximum of 7.83 mm with a mean depth of 3.38 mm. The folds were observed to have two main directions: axial and with an angle of approximately 30 degrees to the axis of the canal. The top of this pine tree-like form points outwards the canal. The folds are filled with mucus that sticks strongly to the tissue. Flushing with saline solution will not remove the mucus.

To determine the possible effects of mucus in the endocervical canal, the transmission and fluorescence of several samples of mucus was measured, and the results are presented in graphical form in Figures 23 and 24. To produce these graphs, small

- 62 -

amounts of mucus were diluted in 10 ml of normal buffered saline solution and placed in a 1 cm pathlength.

As can be seen with reference to Figures 23 and 24, the strongest emission of mucus is at 340 nm emission with an excitation at 280 nm. This does not interfere with the measurements described in this example.

Instrumentation

Referring now to Figure 25, an apparatus is disclosed using a single pixel optical probe. The apparatus includes endocervical probe 11 which incorporates a number of optical fibers including excitation fibers 12, 13 and 14 and collection fiber 16. The excitation fibers are connected to an illumination source which may be, for example, two nitrogen lasers 17,18 (LN300C, Laser Photonics) with a dye module. Other illumination sources, for example a Xenon lamp and filter wheel (disclosed in more detail with reference to Figure 24), may also be used. Other illumination sources may also be acceptable, including, for example, various types of lasers (for example, HeCd or Ag lasers) used with or without dye modules, and various types of so-called white light sources (for example, Xe, Hg, or XeHg lamps) used with filter wheels. This illumination source produces light at frequencies that have been selected for their ability to produce fluorescence in tissue that permits characterization of the tissue. For example light at approximately 337, 380 and 460 nanometers has proven useful. This light is coupled into excitation fibers 12, 13, 14. For coupling, standard Microbench components (Spindler Hoyer) and planoconvex lenses 19 were used. The light coming out of the two dye modules is bandpass filtered by bandpass filters 20 to minimize fluorescence from the dye being coupled into the excitation fibers 12, 13 and 14. Collection fiber 16 collects the fluorescence which is projected through a coupling optics 22 (for example, Microbench, magnification 50/30) into a detector 24, for example an F/3.8 spectrograph (Monospec 18, Thermo Jarrel Ash, Scientific Measurement Systems, Inc.). In the coupling optics 22, longpass filter 23 (for example, color glass filters, Schott) block the reflected excitation light from entering the detector. The spectrograph disperses the light onto an intensified diode array 26. Exemplary diode array 26, electronics

- 63 -

and controller 27 are manufactured by Princeton Instruments. The system also includes gate pulser 28 which is used to control the operation of lasers 17 and 18. Lasers 17 and 18 may be controlled, for example at a 30 Hz repetition rate with a 5 nanosecond pulse duration, but other repetition rates and pulse durations may also be acceptable.

5   The apparatus also includes programmed computer 29 which operates to energize lasers 17 and 18 and to analyze the fluorescence spectra collected by collection fiber 16 in order to characterize the tissue sample under study. The programmed computer 29 is as described in the second example or the third example above.

        Although a single pixel probe was used for this example, a multiple pixel optical
10   probe is also useful. Referring now to Figure 26, an apparatus is disclosed using a multiple pixel optical probe. The apparatus includes a multiple pixel optical probe 21 which incorporates excitation optical fibers 33 and collection optical fibers 34. Excitation optical fibers 33 are connected to receive light from illumination source 35 which may be, for example, a Xenon lamp 26 in combination with a filter wheel 36. Once again, other
15   illumination sources, including for example, the laser source disclosed with reference to Figure 1, would also be acceptable. As with the apparatus of Figure 1, illumination source 35 produces light at frequencies that have been selected for their ability to produce fluorescence in tissue that permits characterization of the tissue.

        Collection fibers 34 from probe 21 are connected to detector 24 which includes, for
20   example, an imaging spectrograph 37 (for example, a Chromex 250 IS), and a CCD array 31 (for example, a thermoelectric cooled CCD Princeton Instruments EV 578x384). The output of detector 24 is applied to computer 32 which is programmed to control illumination source 35 and to analyze the fluorescence spectra collected by collection fibers 34 and detected by detector 24 using, for example, the analysis methods disclosed in the second example or the
25   third example above.

        The transmission and fluorescence of FEP tubing, which is a presently preferred material for use as the housing for the probes, was measured and the results are presented in

- 64 -

Figures 27 and 28. As can be seen with reference to Figures 27 and 28, the fluorescence of the FEP tubing is low. However the autofluorescence of the FEP tubing is about 1/10 of the tissue fluorescence at 337 nm excitation. There is a main emission peak at 400 nm with 320 nm excitation. It was determined that this contribution could be accommodated during the

5   probe calibration procedure discussed below.

Exemplary single and multiple pixel optical probes and various design criteria therefor are described in detail in United States Patent Application Serial No. 08/693,471, Filed August 2, 1996, which hereby is incorporated herein by reference in its entirety.

Method

10   In a clinical application, the method of this example has as its purpose the characterization of epithelial viscus tissue, such as , for example, tissue of the endocervical canal. In general, when applied to the characterization of endocervical tissue, the method has as its purposes to: a) identify lesions extending from the ectocervix into the endocervical canal; b) detect the position of the transformation zone if present inside the endocervical

15   canal; and c) identify squamous lesions with columnar involvement inside the endocervical canal. In general, these purposes are accomplished by measuring fluorescence spectra at spatially resolved locations inside the endocervical canal over a substantially cylindrical area of the interior surface of the tissue of the canal, and using probability-based mathematical models to characterize that tissue as a function of the measured spectra. An accepted method

20   to classify cervical tissues is the new Bethesda system as presented in Wright *et al.* , "Pathology of the Female Genital Tract," 156-177, Springer-Verlag, (1994). In accordance with that system, lesions with HPV and CIN are classified as squamous intraepithelial lesions (SILs) where they may be further separated as high grade SIL (CIN II, CIN III, CIS) and low grade SIL (CIN I, HPV). Normal, metaplastic and non-specific inflammation

25   tissues are classified as non-SILs.

Before beginning a clinical procedure, the measuring apparatus should be calibrated. To calibrate the instrumentation (as shown, for example in Figure 23 and 23), the

- 65 -

background signals are obtained without any excitation which reflects the dark current of the device. This background is stored and is automatically subtracted from any fluorescence measurement. Next, the autofluorescence of the probe is determined, for example, by placing the probe in a brown bottle containing sterile H_2O and measuring fluorescence

5   spectra with the excitation light on. This signal is not subtracted from the tissue fluorescence, however it may be subtracted if desired. In order to confirm calibration, a standard rhodamine solution (OD 0.446725, (= 550 nm, 1 cm pathlength) may be measured. Based on previous clinical work, Rhodamine has been shown to have approximately twice the intensity of squamous cervical tissue fluorescence.

10   During spectral measurement of tissue, if improvement in the signal to noise ratio is desired, the spectra may be accumulated 100 and 200 times, respectively at 380 and 460 nm At 337 nm 50 accumulations have proven sufficient. However, other methods to improve the signal to noise ratio may also be used. For all three wavelengths a different background subtraction file may be used with the corresponding accumulations.

15   During a clinical procedure, it is desired to obtain fluorescence spectra at preferably three excitation wavelengths along the substantially cylindrical surface of the entire endocervical canal with a spatial resolution of approximately 1.5 mm. This may be accomplished by use of either of the apparatus of Figures 23 or 24 with any suitable optical probe, including a single pixel probe, ring probe, line probe, or area probe. During a

20   procedure, the outer housing of the probe is placed and advanced to the internal os of the endocervical canal. Fluorescence measurement are then started. In the case of a single pixel probe, the single measuring pixel is advance both axially and angularly within the housing in order to image a sufficient number of pixels over the substantially cylindrical tissue surface. When using a ring probe, the measuring ring of pixels is advance axially in order to image a

25   sufficient number of pixels over the substantially cylindrical tissue surface. When using a line probe, the measuring line of pixels is incremented angularly in order to image a sufficient number of pixels over the substantially cylindrical tissue surface; for example, four individual measurement may be taken, one each at 12, 3, 6, and 9 o'clock (i.e., every 90°). This procedure takes approximately 3 minutes to complete.

- 66 -

Either before or during a procedure, saline solution may be flushed over the tissue in order possibly to improve measurement accuracy by removing mucus or blood or loose tissue form the measurement site.

5   In general, if the margin of the first specimen at the endocervical side is free of dysplasia or cancer and the second specimen shows no changes it may be assumed that the canal is in a normal condition. If this margin is involved with changes it may be assumed that the first 5 mm of the canal are in an abnormal state. If the margin of the endocervical specimen contains no changes it may be assumed that the margins extend no deeper than 2 cm. If this specimen shows abnormal cells it may be assumed that the measurements in the canal were abnormal even after 5 mm. If the second specimen is marked as metaplasia it may be assumed that the transformation zone is inside the endocervical canal. If the first specimen shows metaplasia the transformation zone is located around the os or on the ectocervix.

15   Figures 29, 30 and 31 present groups of normalized fluorescence intensity spectra obtained *in vivo* from endocervical canals of several different patients using the method and instrumentation of this example. In particular, Figure 29 is a group of normalized fluorescence intensity spectra obtained with 337 nm excitation, Figure 30 is a group of fluorescence intensity spectra obtained using 380 nm excitation, and Figure 31 is a group of normalized fluorescence intensity spectra obtained using 460 nm excitation.

20   Clinical Methods for Performing the Composite Screening
Algorithms of Examples 2, 3 and 4

In a clinical setting, the following exemplary steps are carried out to perform the composite screening algorithm of Examples 2, 3 and 4 above.

25   The instrument is turned on and calibrated. Next, the prior probability that the patient to be measured has SIL is entered. This probability may be derived from statistics from the general population, or may be derived from patient-specific data collected, for example,

- 67 -

from a prior colposcopy. Next, a speculum is inserted and the cervix is observed. Acetic acid may be applied to the cervix, if desired.

The probe is directed to the cervix, ensuring that areas desired for screening will be illuminated. Multiple placements of the probe may be necessary. Using the probe, the cervix

5   is illuminated with excitation at approximately 337 nm, 380 nm and 460 nm. The probe will record resulting fluorescence data.

Data from each spatial location assessed is analyzed to indicate whether the tissue is SIL or not. Analysis steps carried out include the following.

1. Data recorded from each spatial location on the cervix is pre-processed in

10   two ways: normalization, and normalization followed by mean scaling. Similarly pre-processed data obtained at each excitation wavelength are concatenated into a vector for each spatial location assessed.

2. The normalized data vector from each site (Dn') is multiplied by the reduced eigenvector matrix stored in memory (Cn').   Cn' contained only

15   those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 1.

3. The posterior probabilities that a sample is SIL or normal squamous epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal squamous

20   epithelium and SILs and optimal costs of misclassification stored in memory and the entered prior probability are used.

4. The normalized, mean-scaled prediction data vector (Dnm') is multiplied by the reduced eigenvector matrix from normalized, mean-scaled spectral data stored in memory (Cnm'). Cnm' contains only those eigenvectors which

displayed statistically significant differences for samples to be classified by constituent algorithm 2.

5. The posterior probabilities that a sample is SIL or normal columnar epithelium are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for normal columnar epithelium and SILs and optimal costs of misclassification stored in memory and entered prior probabilities are used.

6. Using constituent algorithm 1, sites with a posterior probability of being normal squamous epithelium greater than a threshold value are classified as non-SIL. Remaining sites are classified based on the output of constituent algorithm 2. Using constituent algorithm 2, sample with a posterior probability of being normal columnar epithelium greater than a threshold are classified as non-SIL. The remaining samples are classified as SIL. These tissue classifications may then be displayed in an easily understandable way, for example, by displaying an image of the cervix with the different tissue types displayed as different colors.

To use the composite diagnostic algorithm in clinical practice, the following exemplary steps are carried out.

The instrument is turned on and calibrated. The prior probability that the patient to be measured has SIL and HGSIL is entered. Once again, this probability may be derived from statistics from the general population, or may be derived from patient-specific data collected, for example, from a prior colposcopy. Next, a speculum is inserted and the cervix is observed. Acetic acid may be applied to the cervix, if desired.

The probe is directed to the cervix, ensuring that areas desired for screening will be illuminated. Multiple placements of the probe may be necessary. Using the probe, the cervix

- 69 -

is illuminated with excitation at approximately 337 nm, 380 nm and 460 nm. The probe will record resulting fluorescence data.

Data from each spatial location assessed is analyzed to indicate whether the tissue is HGSIL or not. Analysis steps carried out include:

5   1. Data recorded from each spatial location on the cervix is pre-processed in two ways: normalization, and normalization followed by mean scaling. Similarly pre-processed data obtained at each excitation wavelength are concatenated into a vector for each spatial location assessed.

10   2. The normalized data vector from each site (Dn') is multiplied by the reduced eigenvector matrix stored in memory (Cn'). Cn' contained only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 1.

3. The posterior probabilities that a sample is SIL or normal squamous epithelium are calculated using Bayes theorem. In this calculation, the mean

15   values and standard deviations of the PC scores for normal squamous epithelium and SILs and optimal costs of misclassification stored in memory and the entered prior probability are used.

4. The normalized, mean-scaled prediction data vector (Dnm') is multiplied by the reduced eigenvector matrix from normalized, mean-scaled spectral

20   data stored in memory (Cnm'). Cnm' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 2.

5. The posterior probabilities that a sample is SIL or normal columnar epithelium are calculated using Bayes theorem. In this calculation, the mean

values and standard deviations of the PC scores for normal columnar epithelium and SILs and optimal costs of misclassification stored in memory and entered prior probabilities are used.

6. The normalized prediction data vector (Dn') is multiplied by the reduced eigenvector matrix from normalized spectral data of the calibration set (Cn'). Cn' contains only those eigenvectors which displayed statistically significant differences for samples to be classified by constituent algorithm 3.

7. The posterior probabilities that a sample HGSIL or LGSIL are calculated using Bayes theorem. In this calculation, the mean values and standard deviations of the PC scores for HGSILs and LGSILs and optimal costs of misclassification stored in memory and entered prior probabilities are used.

8. Using constituent algorithm 1, sample with a posterior probability of being normal squamous epithelium greater than a threshold are classified as non-SIL. Remaining samples are classified based on the output of constituent algorithm 2. Using constituent algorithm 2, sample with a posterior probability of being normal columnar epithelium greater than a threshold are classified as non-SIL. Remaining samples are classified based on the output of constituent algorithm 3. Using constituent algorithm 3, samples with a posterior probability of being LGSIL greater than a threshold are classified as LGSIL. The remaining samples are classified as HGSIL. These tissue classifications may then be displayed in an easily understandable way, for example, by displaying an image of the cervix with the different tissue types displayed as different colors.

The previous examples and clinical methods are included to demonstrate specific embodiments. It will be appreciated by those of skill in the art that the techniques disclosed in the examples and the clinical methods represent techniques discovered by the inventors to function well in the practice of the technology, and thus can be considered to constitute

- 71 -

specific modes for its practice. Those of skill in the art will also appreciate in light of the present disclosure, that variations and modifications of the methods and apparatus disclosed herein are possible, and that practical alternatives to and equivalents of the various elements of the methods and apparatus may be practiced without departing from the scope and spirit

5   of the invention. Accordingly, the description and applications as set forth herein are illustrative and are not intended to limit the scope of the invention, which is defined in the following claims.

A: 231018(4Y9601!.DOC)